



Amazon EMR-Versionshinweise

# Amazon EMR



# Amazon EMR: Amazon EMR-Versionshinweise

Copyright © 2024 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

Die Handelsmarken und Handelsaufmachung von Amazon dürfen nicht in einer Weise in Verbindung mit nicht von Amazon stammenden Produkten oder Services verwendet werden, durch die Kunden irregeführt werden könnten oder Amazon in schlechtem Licht dargestellt oder diskreditiert werden könnte. Alle anderen Marken, die nicht im Besitz von Amazon sind, gehören den jeweiligen Besitzern, die möglicherweise mit Amazon verbunden sind oder von Amazon gesponsert werden.

---

# Table of Contents

Informationen über Amazon-EMR-Versionen .....	1
Amazon-EMR-7.x-Versionen .....	2
Anwendungsversionen in Amazon-EMR-7.x-Versionen .....	3
emr-7.1.0 .....	3
emr-7.0.0 .....	30
Amazon-EMR-6.x-Versionen .....	66
Anwendungsversionen in Amazon-EMR-6.x-Versionen .....	68
emr-6.15.0 .....	68
emr-6.14.0 .....	114
emr-6.13.0 .....	160
emr-6.12.0 .....	213
emr-6.11.1 .....	271
emr-6.11.0 .....	325
emr-6.10.1 .....	362
emr-6.10.0 .....	417
emr-6.9.1 .....	457
emr-6.9.0 .....	509
emr-6.8.1 .....	555
emr-6.8.0 .....	606
emr-6.7.0 .....	653
emr-6.6.0 .....	716
emr-6.5.0 .....	782
emr-6.4.0 .....	810
emr-6.3.1 .....	845
emr-6.3.0 .....	872
emr-6.2.1 .....	906
emr-6.2.0 .....	934
emr-6.1.1 .....	968
emr-6.1.0 .....	989
emr-6.0.1 .....	1015
emr-6.0.0 .....	1033
Amazon-EMR-5.x-Versionen .....	1055
Anwendungsversionen in Amazon-EMR-5.x-Versionen (PNG) .....	1059
emr-5.36.2 .....	1059

---

emr-5.36.1 .....	1085
emr-5.36.0 .....	1139
emr-5.35.0 .....	1177
emr-5.34.0 .....	1205
emr-5.33.1 .....	1231
emr-5.33.0 .....	1261
emr-5.32.1 .....	1284
emr-5.32.0 .....	1308
emr-5.31.1 .....	1337
emr-5.31.0 .....	1357
emr-5.30.2 .....	1380
emr-5.30.1 .....	1400
emr-5.30.0 .....	1422
emr-5.29.0 .....	1446
emr-5.28.1 .....	1465
emr-5.28.0 .....	1484
emr-5.27.1 .....	1504
emr-5.27.0 .....	1522
emr-5.26.0 .....	1541
emr-5.25.0 .....	1561
emr-5.24.1 .....	1580
emr-5.24.0 .....	1598
emr-5.23.1 .....	1617
emr-5.23.0 .....	1634
emr-5.22.0 .....	1653
emr-5.21.2 .....	1672
emr-5.21.1 .....	1689
emr-5.21.0 .....	1707
emr-5.20.1 .....	1726
emr-5.20.0 .....	1743
emr-5.19.1 .....	1763
emr-5.19.0 .....	1780
emr-5.18.1 .....	1798
emr-5.18.0 .....	1815
emr-5.17.2 .....	1832
emr-5.17.1 .....	1849



---

emr-5.17.0 .....	1866
emr-5.16.1 .....	1883
emr-5.16.0 .....	1900
emr-5.15.1 .....	1917
emr-5.15.0 .....	1933
emr-5.14.2 .....	1950
emr-5.14.1 .....	1966
emr-5.14.0 .....	1982
emr-5.13.1 .....	2000
emr-5.13.0 .....	2015
emr-5.12.3 .....	2031
emr-5.12.2 .....	2047
emr-5.12.1 .....	2063
emr-5.12.0 .....	2078
emr-5.11.4 .....	2095
emr-5.11.3 .....	2111
emr-5.11.2 .....	2126
emr-5.11.1 .....	2142
emr-5.11.0 .....	2157
emr-5.10.1 .....	2174
emr-5.10.0 .....	2189
emr-5.9.1 .....	2206
emr-5.9.0 .....	2221
emr-5.8.3 .....	2238
emr-5.8.2 .....	2253
emr-5.8.1 .....	2268
emr-5.8.0 .....	2283
emr-5.7.1 .....	2299
emr-5.7.0 .....	2314
emr-5.6.1 .....	2330
emr-5.6.0 .....	2345
emr-5.5.4 .....	2360
emr-5.5.3 .....	2375
emr-5.5.2 .....	2390
emr-5.5.1 .....	2405
emr-5.5.0 .....	2420

emr-5.4.1 .....	2436
emr-5.4.0 .....	2451
emr-5.3.2 .....	2466
emr-5.3.1 .....	2481
emr-5.3.0 .....	2496
emr-5.2.3 .....	2511
emr-5.2.2 .....	2526
emr-5.2.1 .....	2541
emr-5.2.0 .....	2556
emr-5.1.1 .....	2571
emr-5.1.0 .....	2586
emr-5.0.3 .....	2600
emr-5.0.2 .....	2615
emr-5.0.1 .....	2629
emr-5.0.0 .....	2643
Amazon-EMR-4.x-Versionen .....	2658
Anwendungsversionen in Amazon-EMR-4.x-Versionen .....	2659
Versionsunterschiede .....	2660
emr-4.9.6 .....	2707
emr-4.9.5 .....	2722
emr-4.9.4 .....	2736
emr-4.9.3 .....	2750
emr-4.9.2 .....	2764
emr-4.9.1 .....	2779
emr-4.8.5 .....	2793
emr-4.8.4 .....	2808
emr-4.8.3 .....	2822
emr-4.8.2 .....	2836
emr-4.8.1 .....	2851
emr-4.8.0 .....	2865
emr-4.7.4 .....	2879
emr-4.7.3 .....	2893
emr-4.7.2 .....	2906
emr-4.7.1 .....	2920
emr-4.7.0 .....	2933
emr-4.6.1 .....	2947

---

emr-4.6.0 .....	2960
emr-4.5.0 .....	2973
emr-4.4.0 .....	2984
emr-4.3.0 .....	2996
emr-4.2.0 .....	3007
emr-4.1.0 .....	3018
emr-4.0.0 .....	3027
AMI-Versionen 2.x und 3.x .....	3034
Erstellen eines Clusters .....	3035
Installieren von Anwendungen .....	3038
Anpassen von Konfigurationen .....	3038
Hive .....	3045
HBase .....	3056
Pig .....	3070
Spark .....	3076
S3 DistCp .....	3080
Neuerungen? .....	3083
Amazon EMR 7.1.0 .....	3083
Amazon EMR 6.15.0 .....	3086
Amazon EMR 5.36,2 .....	3104
SigV4-Kompatibilität .....	3106
Ansatz zur Minderung von CVE-2021-44228 .....	3107
Amazon-EMR-Bootstrap-Aktionslösung für Log4J CVE-2021-44228 und CVE-2021-45046 .....	3108
Häufig gestellte Fragen .....	3115
Archiv .....	3118
Version 6.14.0 .....	3118
Version 6.13.0 .....	3136
Version 6.12.0 .....	3159
Version 6.11.1 .....	3187
Version 6.11.0 .....	3212
Version 6.10.0 .....	3222
Version 6.9.0 .....	3234
Version 6.8.0 .....	3252
Version 6.7.0 .....	3272
Version 6.6.0 .....	3309
Version 5.35.0 .....	3349

---

Version 5.34.0 .....	3353
Version 6.5.0 .....	3356
Version 6.4.0 .....	3358
Version 5.32.0 .....	3366
Version 6.2.0 .....	3372
Version 5.31.0 .....	3380
Version 6.1.0 .....	3386
Version 6.0.0 .....	3392
Version 5.30.1 .....	3398
Version 5.30.0 .....	3403
Version 5.29.0 .....	3408
Version 5.28.1 .....	3410
Version 5.28.0 .....	3411
Veröffentlichung 5.27.0 .....	3413
Version 5.26.0 .....	3416
Version 5.25.0 .....	3419
Version 5.24.1 .....	3422
Version 5.24.0 .....	3423
Version 5.23.0 .....	3426
Version 5.22.0 .....	3428
Version 5.21.1 .....	3431
Version 5.21.0 .....	3432
Version 5.20.0 .....	3435
Version 5.19.0 .....	3439
Version 5.18.0 .....	3440
Version 5.17.1 .....	3441
Version 5.17.0 .....	3442
Version 5.16.0 .....	3443
Version 5.15.0 .....	3445
Version 5.14.1 .....	3446
Version 5.14.0 .....	3446
Version 5.13.0 .....	3448
Version 5.12.2 .....	3449
Version 5.12.1 .....	3449
Version 5.12.0 .....	3449
Version 5.11.3 .....	3451

---

Version 5.11.2 .....	3452
Version 5.11.1 .....	3452
Version 5.11.0 .....	3453
Version 5.10.0 .....	3454
Version 5.9.0 .....	3456
Version 5.8.2 .....	3458
Version 5.8.1 .....	3458
Version 5.8.0 .....	3458
Version 5.7.0 .....	3460
Version 5.6.0 .....	3461
Version 5.5.3 .....	3462
Version 5.5.2 .....	3462
Version 5.5.1 .....	3463
Version 5.5.0 .....	3463
Version 5.4.0 .....	3464
Version 5.3.1 .....	3465
Version 5.3.0 .....	3466
Version 5.2.2 .....	3466
Version 5.2.1 .....	3467
Version 5.2.0 .....	3468
Version 5.1.0 .....	3469
Version 5.0.3 .....	3469
Version 5.0.0 .....	3470
Version 4.9.5 .....	3471
Version 4.9.4 .....	3472
Version 4.9.3 .....	3472
Version 4.9.2 .....	3472
Version 4.9.1 .....	3473
Version 4.8.4 .....	3473
Version 4.8.3 .....	3474
Version 4.8.2 .....	3475
Version 4.8.0 .....	3475
Version 4.7.2 .....	3476
Version 4.7.1 .....	3477
Version 4.7.0 .....	3477
Version 4.6.0 .....	3479

Version 4.5.0 .....	3481
Version 4.4.0 .....	3481
Version 4.3.0 .....	3483
Version 4.2.0 .....	3485
Anwendungen konfigurieren .....	3487
Anwendungen bei der Clustererstellung konfigurieren .....	3490
Geben Sie beim Erstellen eines Clusters eine Konfiguration in der Konsole an .....	3490
Geben Sie AWS CLI beim Erstellen eines Clusters eine Konfiguration an, indem Sie .....	3490
Beim Erstellen eines Clusters eine Konfiguration mit dem Java SDK angeben .....	3491
Eine Instance-Gruppe in einem laufenden Cluster neu konfigurieren .....	3492
Überlegungen bei der Neukonfiguration einer Instance-Gruppe .....	3493
Eine Instance-Gruppe in der Konsole neu konfigurieren .....	3496
Eine Instance-Gruppe mithilfe der CLI rekonfigurieren .....	3497
Eine Instance-Gruppe mithilfe des Java SDK neu konfigurieren .....	3502
Fehlerbehebung .....	3503
Vertrauliche Konfigurationsdaten in AWS Secrets Manager speichern .....	3506
Ein Secret erstellen .....	3506
Gewähren Sie Amazon EMR Zugriff, um das Geheimnis abzurufen .....	3506
Das Secret in einer Konfigurationsklassifizierung verwenden .....	3507
Den Secret-Wert aktualisieren .....	3508
Anwendungen für die Verwendung einer bestimmten Java Virtual Machine konfigurieren .....	3509
Überlegungen .....	3509
Die JVM überschreiben .....	3512
Service-Ports .....	3514
Anwendungsbenutzer .....	3516
Überprüfen von Abhängigkeiten mit dem Artefakt-Repository .....	3517
EMR File System (EMRFS) .....	3520
Consistent view .....	3522
Die konsistente Ansicht aktivieren .....	3526
Verstehen, wie mit der konsistenten Ansicht von EMRFS Objekte in Amazon S3 nachverfolgt werden .....	3528
Logik für Wiederholversuche .....	3529
Konsistente Ansicht der Metadaten in EMRFS .....	3531
Konsistenzbenachrichtigungen für CloudWatch und Amazon SQS konfigurieren .....	3534
Konfigurieren der konsistenten Ansicht .....	3536
Die EMRFS-CLI-Befehlsreferenz .....	3542

Zugriff auf EMRFS-Daten in Amazon S3 genehmigen .....	3553
Erstellen eines benutzerdefinierten Anmeldeinformationsanbieters für EMRFS-Daten in Amazon S3 .....	3554
Verwaltung des Standardendpunkts AWS Security Token Service .....	3555
Angeben der Amazon-S3-Verschlüsselung mit EMRFS-Eigenschaften .....	3557
Wird AWS KMS keys für die EMRFS-Verschlüsselung verwendet .....	3558
Serverseitige Verschlüsselung im Amazon S3 .....	3559
Clientseitige Verschlüsselung für Amazon S3 .....	3562
CloudWatch Amazon-Vertreter .....	3571
Erstellen eines -Clusters .....	3573
Erforderliche Berechtigungen .....	3574
Erforderlicher Endpunkt .....	3574
Den Cluster erstellen .....	3574
Standardmetriken .....	3575
Konfiguration .....	3578
Amazon EMR 7.1.0 .....	3579
Amazon EMR 7.0.0 .....	3613
Überlegungen .....	3619
Verlauf .....	3619
Delta Lake .....	3621
Einführung .....	3622
Delta-Lake-Cluster verwenden .....	3622
Delta Lake mit Flink .....	3622
Delta Lake mit Trino .....	3627
Delta Lake mit Spark .....	3629
Delta Lake mit Spark und Glue .....	3634
Überlegungen .....	3635
Verlauf .....	3635
Flink .....	3637
Erstellen eines Clusters mit Flink .....	3640
Flink konfigurieren .....	3641
Hive und Glue .....	3641
Konfigurationsdatei .....	3643
Mehrere Primärknoten .....	3644
Speicherprozessgröße .....	3645
Protokollausgabedateigröße .....	3646

---

Java 11 .....	3647
Flink-Aufträge .....	3652
Eine Flink-YARN-Anwendung als Schritt auf einem Cluster mit langer Laufzeit starten .....	3653
Senden Sie Ihre Arbeit an eine bestehende Flink-Anwendung auf einem Cluster mit langer Laufzeit .....	3654
Senden eines kurzlebigen Flink-Auftrags .....	3656
Flink-Scala-Shell .....	3658
Flink-Benutzeroberfläche .....	3659
Flink Autoscaler .....	3660
Übersicht .....	3660
Überlegungen .....	3660
Autoscaler aktivieren .....	3661
Konfigurationen .....	3662
Neustartzeiten optimieren .....	3668
Aufgabenlokale Wiederherstellung .....	3668
Inkrementelle Prüfpunkte .....	3669
Differenzierte Wiederherstellung .....	3670
Kombinierter Neustart .....	3671
Flink mit Zeppelin .....	3672
Einführung .....	3672
Voraussetzungen .....	3673
Zeppelin-Flink auf einem EMR-Cluster konfigurieren .....	3673
Führen Sie Flink-Aufträge mit Zeppelin-Flink auf einem EMR-Cluster aus .....	3675
Versionsverlauf von Flink .....	3680
Ganglia .....	3718
Erstellen eines Clusters mit Ganglia .....	3720
Anzeigen von Ganglia-Metriken .....	3721
Hadoop- und Spark-Metriken in Ganglia .....	3722
Versionsverlauf von Ganglia .....	3723
Hadoop .....	3769
Konfigurieren von Hadoop .....	3771
Aufgabenkonfiguration .....	3772
Hadoop-Daemon-Konfigurationseinstellungen .....	4133
HDFS-Konfiguration .....	4403
Transparente Verschlüsselung in HDFS auf Amazon EMR .....	4405
Konfigurieren der transparenten HDFS-Verschlüsselung .....	4406



Überlegungen zur transparenten HDFS-Verschlüsselung .....	4409
Hadoop-Schlüsselverwaltungsserver (KMS) .....	4409
Transparente HDFS-Verschlüsselung auf EMR-Clustern mit mehreren Hauptknoten .....	4413
Erstellen oder Ausführen einer Hadoop-Anwendung .....	4415
Erstellen von Binärdateien mit Amazon EMR .....	4415
Daten per Streaming verarbeiten .....	4418
Daten mit einer benutzerdefinierten JAR-Datei verarbeiten .....	4423
Aktivieren Sie die Erkennung ungleichmäßiger Speicherzugriffe für YARN-Container .....	4427
Hadoop-Versionsverlauf .....	4429
Hadoop-Versionshinweise nach Version .....	4491
HBase .....	4495
Erstellen eines Clusters mit HBase .....	4499
Erstellen eines Clusters mit HBase über die Konsole .....	4499
Erstellen eines Clusters mit HBase unter Verwendung des AWS CLI .....	4499
HBase auf Amazon S3 (Amazon-S3-Speichermodus) .....	4501
Aktivieren von HBase auf Amazon S3 .....	4502
Verwenden eines Lesereplikat-Clusters .....	4503
Persistentes HFile-Tracking .....	4505
Betriebliche Überlegungen .....	4507
Write-Ahead-Protokolle (WAL) für Amazon EMR .....	4511
WAL-Workspaces .....	4512
Erforderliche Berechtigungen .....	4513
Aktivieren von WAL .....	4514
Wiederherstellen aus WAL .....	4517
Sicherheitskonfigurationen .....	4518
Verwenden von AWS PrivateLink .....	4519
WAL-Preise und -Metriken .....	4520
Markieren von WAL-Workspaces .....	4522
Überlegungen und Verfügbarkeit .....	4523
EMRWAL-CLI-Referenz .....	4524
Verwenden der HBase-Shell .....	4528
Erstellen einer Tabelle .....	4528
Eingeben eines Werts .....	4529
Abrufen eines Werts .....	4529
Löschen einer Tabelle .....	4529
Zugreifen auf HBase-Tabellen mit Hive .....	4529

Verwenden von HBase-Snapshots .....	4531
Erstellen eines Snapshots mit einer Tabelle .....	4531
Löschen eines Snapshots .....	4532
Anzeigen von Snapshot-Informationen .....	4532
Einen Snapshot nach Amazon S3 exportieren .....	4532
Importieren eines Snapshots von Amazon S3 .....	4533
Wiederherstellen einer Tabelle von Snapshots innerhalb der HBase-Shell .....	4534
Konfigurieren von HBase .....	4535
Änderungen an der Arbeitsspeicherzuordnung in YARN .....	4536
HBase-Port-Nummern .....	4536
Zu optimierende HBase-Standorteinstellungen .....	4537
Anzeigen der HBase-Benutzeroberfläche .....	4539
Anzeigen der HBase-Protokolldateien .....	4541
Überwachen von HBase mit Ganglia .....	4542
Migrieren von vorherigen HBase-Versionen .....	4544
Versionsverlauf von HBase .....	4544
HCatalog .....	4616
Erstellen eines Clusters mit HCatalog .....	4617
Verwenden von HCatalog .....	4618
Deaktivieren der Direct-Write-Funktion bei der Verwendung von HCatalog HStorer .....	4619
Erstellen einer Tabelle mit der HCat CLI und Verwenden der entsprechenden Daten in Pig .....	4619
Zugreifen auf die Tabelle mit Spark SQL .....	4620
Beispiel: Erstellen einer HCatalog-Tabelle und Ausführen eines Schreibvorgangs mit Pig .....	4622
Versionsverlauf von HCatalog .....	4623
Hive .....	4691
Unterschiede und Überlegungen für Hive auf Amazon EMR .....	4694
Unterschiede zwischen Apache Hive in Amazon EMR und Apache Hive .....	4694
Hive-Unterschiede zwischen Amazon-EMR-Version 4.x und 5.x .....	4695
Zusätzliche Features von Hive in Amazon EMR .....	4696
Einen externen Metastore für Hive konfigurieren .....	4701
Verwenden des AWS Glue-Datenkatalogs als Metastore für Hive .....	4702
Verwenden einer externen MySQL-Datenbank oder von Amazon Aurora; .....	4710
Den Hive-JDBC-Treiber verwenden .....	4713
Verbessern der Hive-Leistung .....	4715
Aktivierung des optimierten Hive EMRFS S3 Committers .....	4715
Verwendung von S3 Select .....	4717

MSCK-Optimierung .....	4719
Verwenden von Hive LLAP .....	4720
LLAP aktivieren .....	4721
Starten Sie LLAP auf Ihrem Cluster .....	4722
Prüfen des LLAP-Status .....	4723
LLAP starten oder beenden .....	4723
Ändern Sie die Größe der Anzahl der LLAP-Daemons .....	4723
Verschlüsselung im Hive .....	4723
Modulare Parquet-Verschlüsselung in Hive .....	4724
Verschlüsselung während der Übertragung HS2 .....	4728
Versionsverlauf von Hive .....	4729
Hive-Versionshinweise nach Version .....	4799
Hudi .....	4884
Funktionsweise von Hudi .....	4886
Grundlegendes zu den Speichertypen von Datensätzen: „Beim Schreiben kopieren“ gegenüber „beim Lesen zusammenführen“ .....	4886
Registrieren eines Hudi-Datensatzes bei Ihrem Metastore .....	4888
Überlegungen und Einschränkungen .....	4888
Erstellen eines Cluster mit installierter Hudi-Anwendung .....	4890
Arbeiten mit einem Hudi-Datensatz .....	4891
Initialisieren Sie eine Spark-Sitzung für Hudi .....	4895
Schreiben Sie in einen Hudi-Datensatz .....	4896
Upsert Daten .....	4901
Einen Datensatz löschen .....	4902
Aus einem Hudi-Datensatz lesen .....	4904
Die Hudi-CLI verwenden .....	4906
Hudi-Versionsverlauf .....	4907
Hue .....	4910
Hue-Versionen .....	4910
Unterstützte und nicht unterstützte Features von Hue in Amazon EMR .....	4912
Überlegungen .....	4913
Leistung bei großen Hue-Metadatentabellen .....	4914
Inkompatibilität zwischen Hue-Versionen .....	4915
Verbindung zur Hue-Webbenutzeroberfläche herstellen .....	4915
Verwenden von Hue mit einer Remote-Datenbank in Amazon RDS .....	4916
Fehlerbehebung .....	4918

Erweiterte Konfigurationen für Hue .....	4919
Konfigurieren von Hue für LDAP-Benutzer .....	4919
Hue-Versionsverlauf .....	4922
Iceberg .....	4984
Funktionsweise von Iceberg .....	4985
Einen Cluster mit Iceberg verwenden .....	4987
Verwenden Sie einen Iceberg-Cluster mit Spark .....	4987
Verwenden Sie einen Iceberg-Cluster mit Trino .....	4992
Verwenden Sie einen Iceberg-Cluster mit Flink .....	4994
Verwenden Sie einen Iceberg-Cluster mit Hive .....	4999
Überlegungen und Einschränkungen .....	5002
Überlegungen zur Verwendung von Iceberg mit Spark .....	5002
Überlegungen zur Verwendung von Iceberg mit Trino .....	5003
Überlegungen zur Verwendung von Iceberg mit Flink .....	5003
Überlegungen zur Verwendung von Iceberg mit Hive .....	5003
Versionsverlauf von Iceberg .....	5004
Iceberg-Versionshinweise nach Version .....	5005
Jupyter Notebook .....	5007
EMR Studio .....	5007
EMR Notebook .....	5007
JupyterHub .....	5008
Erstellen Sie einen Cluster mit JupyterHub .....	5012
Überlegungen bei der Verwendung JupyterHub auf Amazon EMR .....	5014
Konfiguration JupyterHub .....	5015
Konfigurieren von Persistenz für Notebooks in Amazon S3 .....	5016
Eine Verbindung mit dem Hauptknoten und Notebook-Servern herstellen .....	5017
JupyterHub Konfiguration und Verwaltung .....	5018
Hinzufügen von Jupyter-Notebook-Benutzern und Administratoren .....	5020
Installieren von zusätzlicher Kerneln und Bibliotheken .....	5032
JupyterHub Versionsverlauf .....	5037
Livy .....	5073
Aktivieren von HTTPS .....	5075
Livy-Versionsverlauf .....	5076
MXNet .....	5120
MXNet-Versionsverlauf .....	5122
Oozie .....	5151

Oozie mit einer Remote-Datenbank in Amazon RDS verwenden .....	5154
Die Java-Version für Oozie konfigurieren .....	5156
Versionsverlauf von Oozie .....	5157
Versionshinweise zu Oozie nach Version .....	5214
Phoenix .....	5215
Einen Cluster mit Phoenix erstellen .....	5218
Anpassen von Phoenix-Konfigurationen .....	5219
Phoenix-Clients .....	5220
Versionsverlauf von Phoenix .....	5224
Pig .....	5297
Übermitteln von Pig-Aufträgen .....	5299
Pig-Aufträge über die Amazon-EMR-Konsole senden .....	5300
Reichen Sie Arbeiten von Pig ein, indem Sie AWS CLI .....	5301
Benutzerdefinierte Funktionen in Pig aufrufen .....	5303
JAR-Dateien in Pig aufrufen .....	5303
Aufrufen von Python/Jython-Skripts in Pig .....	5303
Pig-Versionsverlauf .....	5305
Presto und Trino .....	5370
Presto mit dem AWS Glue-Datenkatalog verwenden .....	5373
AWS Glue Data Catalog als Metastore angeben .....	5374
IAM-Berechtigungen .....	4706
Überlegungen zur Verwendung von AWS Glue Data Catalog .....	5379
Verwenden von S3 Select-Pushdown .....	5380
Ist S3-Select-Pushdown das Richtige für meine Anwendung? .....	5380
Überlegungen und Einschränkungen .....	5381
Aktivieren von S3-Select-Pushdown für Presto oder Trino .....	5381
Hinzufügen von Datenbank-Konnektoren .....	5382
Verwenden von SSL/TLS und LDAPS .....	5383
Verwenden der LDAP-Authentifizierung .....	5385
Den strikten Presto-Modus aktivieren .....	5393
Überlegungen .....	5395
Umgang mit dem Verlust von Spot Instances in Presto .....	5395
Fehlertolerante Ausführung .....	5398
Konfiguration .....	5398
Austausch-Manager .....	5399
Überlegungen und Einschränkungen .....	5400

Auto Scaling von Presto mit ordnungsgemäßer Stilllegung verwenden .....	5400
Überlegungen zu Presto in Amazon EMR .....	5402
In der Befehlszeile von Presto ausführbare Datei .....	5402
Nicht konfigurierbare Presto-Bereitstellungseigenschaften .....	5402
Installation von PrestoDB und Trino .....	5404
EMRFS- und PrestoS3-Konfiguration FileSystem .....	5404
Standardeinstellung für Endbenutzer-Identitätswechsel .....	5405
Standardport für die Presto-Webschnittstelle .....	5405
Problem bei der Hive-Bucket-Ausführung in einigen Versionen .....	5405
Presto-Versionsverlauf .....	5406
Trino (PrestoSQL) Versionshinweise nach Version .....	5471
Spark .....	5474
Einen Spark-Cluster erstellen .....	5478
Spark-Anwendungen mit Docker mithilfe von Amazon EMR 6.x ausführen .....	5481
Überlegungen .....	5481
Ein Docker-Image erstellen .....	5482
Verwenden von Docker-Images von Amazon ECR .....	5483
Verwenden Sie den AWS Glue-Datenkatalog als Metastore für Spark SQL .....	5489
AWS Glue Data Catalog als Metastore angeben .....	5490
IAM-Berechtigungen .....	4706
Überlegungen .....	4708
Konfigurieren von Spark .....	5495
Spark-Standardwerte .....	5496
Garbage Collection einrichten .....	5498
maximizeResourceAllocation .....	5499
Verhalten bei Außerbetriebnahme .....	5501
ThriftServer Spark-Umgebungsvariable .....	5505
Ändern von Spark-Standardeinstellungen .....	5506
Migration von Apache Log4j 1.x zu Log4j 2.x .....	5508
Die Leistung von Spark optimieren .....	5509
Adaptive Abfrageausführung .....	5509
Dynamische Partitionsbereinigung .....	5512
Abflachen skalarer Unterabfragen .....	5514
DISTINCT vor INTERSECT .....	5515
Bloomfilter für Joins .....	5516
Optimierte Join-Neuanordnung .....	5516

Caching von Ergebnisfragmenten .....	5517
Caching von Ergebnisfragmenten in Spark aktivieren .....	5518
Überlegungen .....	5519
RAPIDS Accelerator verwenden .....	5521
Instance-Typen auswählen .....	5521
App-Konfigurationen einrichten .....	5521
Eine Bootstrap-Aktion hinzufügen .....	5529
Ihren Cluster starten .....	5530
Zugriff auf die Spark-Shell .....	5530
Verwenden Sie Amazon SageMaker Spark für maschinelles Lernen .....	5532
Eine Spark-Anwendung schreiben .....	5532
Scala .....	5532
Java .....	5533
Python .....	5534
Spark-Leistung mit S3 verbessern .....	5536
S3 Select verwenden .....	5536
EMRFS-S3-optimierter Committer .....	5541
Verwenden Sie das EMRFS-S3-optimierte Commit-Protokoll .....	5548
S3-Anforderungen wiederholen .....	5555
Einen Spark-Schritt hinzufügen .....	5558
Überschreiben der standardmäßigen Spark-Konfigurationseinstellungen .....	5562
Anzeigen des Spark-Anwendungsverlaufs .....	5562
Zugreifen auf Web-Benutzerschnittstellen in Spark .....	5563
Verwenden von Spark mit Amazon Kinesis Data Streams .....	5563
Spark auf Amazon Redshift verwenden .....	5564
Eine Spark-Anwendung starten .....	5564
Bei Amazon Redshift authentifizieren .....	5565
In Amazon Redshift schreiben und lesen .....	5568
Überlegungen .....	5569
Spark-Versionsverlauf .....	5571
Sqoop .....	5634
Sqoop-Versionen .....	5634
Überlegungen zu Sqoop auf Amazon EMR .....	5636
Sqoop mit der HCatalog-Integration verwenden .....	5636
Sqoop JDBC- und Datenbank-Support .....	5637
Schützen Ihres Passworts .....	5638

---

Versionsverlauf von Sqoop .....	5639
TensorFlow .....	5692
TensorFlow Builds nach Amazon EC2 EC2-Instanztyp .....	5694
Sicherheit .....	5694
Verwenden TensorBoard .....	5694
TensorFlow Verlauf der Veröffentlichung .....	5695
Tez .....	5718
Erstellen eines Clusters mit Tez .....	5720
Konfigurieren von Tez .....	5721
Beispielkonfiguration .....	5721
Asynchrone Öffnung von Tez-Splits .....	5722
Tez-Webbenutzeroberfläche .....	5724
Timeline-Server .....	5725
Tez-Versionsverlauf .....	5725
Tez-Versionshinweise nach Version .....	5769
Zeppelin .....	5776
Überlegungen zur Verwendung von Zeppelin in Amazon EMR .....	5778
Zeppelin-Versionshistorie .....	5779
ZooKeeper .....	5837
ZooKeeper Verlauf der Veröffentlichung .....	5839
Konnektoren und Dienstprogramme .....	5878
Tabellen in DynamoDB exportieren, importieren, abfragen und verknüpfen .....	5878
Eine Hive-Tabelle einrichten, um Hive-Befehle auszuführen .....	5880
Hive-Befehlsbeispiele für das Exportieren, Importieren und Abfragen von Daten .....	5890
Optimierung der Leistung .....	5900
Kinesis .....	5904
Wie kann ich die Integration von Amazon EMR und Amazon Kinesis verwenden? .....	5904
Analyse von Amazon-Kinesis-Streams mit Checkpointing .....	5904
Leistungsaspekte .....	5906
Die Amazon-Kinesis-Analyse mit Amazon EMR planen .....	5906
Migration des Spark-Kinesis-Konnektors zu SDK 2.x für Amazon EMR 7.0 .....	5906
DistCp S3 (s3-dist-cp) .....	5915
DistCp S3-Optionen .....	5916
S3 DistCp als Schritt in einem Cluster hinzufügen .....	5924
Aufräumen nach fehlgeschlagenen DistCp S3-Jobs .....	5926
Befehle und Skripte auf einem Cluster ausführen .....	5928



---

Reichen Sie einen benutzerdefinierten JAR-Schritt ein, um ein Skript oder einen Befehl auszuführen .....	5929
Andere Verwendungsmöglichkeiten von <code>command-runner.jar</code> .....	5930
AWS-Glossar .....	5932
.....	5933

# Informationen über Amazon-EMR-Versionen

Eine Amazon-EMR-Version ist eine Gruppe von Open-Source-Anwendungen aus dem Big-Data-Ökosystem. Jede Version besteht aus verschiedenen Big-Data-Anwendungen, Komponenten und Features, die Amazon EMR installieren und konfigurieren soll, wenn Sie einen Cluster erstellen. Anwendungen werden mithilfe eines Systems verpackt, das auf [Apache](#) basiert BigTop, einem Open-Source-Projekt, das mit dem Hadoop-Ökosystem verbunden ist. Dieses Handbuch enthält Informationen für Anwendungen, die in Amazon-EMR-Versionen enthalten sind.

Weitere Informationen zu den ersten Schritten und der Arbeit mit Amazon EMR finden Sie im [Verwaltungshandbuch für Amazon EMR](#).

Wenn Sie einen Cluster starten, können Sie aus verschiedenen Versionen von Amazon EMR wählen. Auf diese Weise können Sie Anwendungsversionen testen und verwenden, die zu den Kompatibilitätsanforderungen passen. Sie geben die Versionsnummer zusammen mit dem Release-Label an. Versionsbezeichnungen haben die Form `emr-x.x.x`. z. B. `emr-7.1.0`.

Ab Amazon EMR 5.18.0 können Sie das Amazon-EMR-Artefakt-Repository verwenden, um Ihren Auftragscode anhand der genauen Versionen von Bibliotheken und Abhängigkeiten zu erstellen, die mit bestimmten Amazon-EMR-Versionen verfügbar sind. Weitere Informationen finden Sie unter [Überprüfen von Abhängigkeiten mithilfe des Amazon-EMR-Artefakt-Repositorys](#).

Wenn Sie Updates erhalten möchten, wenn eine neue Amazon-EMR-Version verfügbar ist, abonnieren Sie das [RSS-Feed für Amazon-EMR-Versionshinweise](#).

Details zu neuesten Versionen, einschließlich Anwendungsversionen, Versionshinweisen, Komponenten und Konfigurationsklassifizierungen der Amazon-EMR-Serien 7.x, 6.x und 5.x:

- [Amazon EMR Version 7.1.0](#)
- [Amazon-EMR-Version 6.15.0](#)
- [Amazon EMR Version 5.36.2](#)

## Note

Neue Amazon-EMR-Versionen werden über einen Zeitraum von mehreren Tagen in verschiedenen Regionen verfügbar gemacht, beginnend mit der ersten Region am

ursprünglichen Veröffentlichungsdatum. Die neueste Version ist in diesem Zeitraum möglicherweise nicht in Ihrer Region verfügbar.

Versionshinweise für die neuesten Amazon-EMR-Versionen und ein Verlauf aller Versionen:

- [Neuerungen?](#)
- [Amazon-EMR-Archiv mit Versionshinweisen](#)

Ein umfassender Verlauf von Anwendungsversionen in jeder Amazon-EMR-Version:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

Details für jede Amazon-EMR-Version und Unterschiede zwischen Versionsreihen, wenn zutreffend:

- [Amazon-EMR-7.x-Versionen](#)
- [Amazon-EMR-6.x-Versionen](#)
- [Amazon-EMR-5.x-Versionen](#)
- [Amazon-EMR-4.x-Versionen](#)
- [AMI-Versionen 2.x und 3.x von Amazon EMR](#)

## Amazon-EMR-7.x-Versionen

Dieser Abschnitt enthält Anwendungsversionen, Versionshinweise, Komponentenversionen und Konfigurationsklassifizierungen, die in jeder Amazon-EMR-7.x-Version verfügbar ist.

Wenn Sie einen Cluster starten, können Sie aus verschiedenen Versionen von Amazon EMR wählen. Auf diese Weise können Sie Anwendungsversionen testen und verwenden, die zu den Kompatibilitätsanforderungen passen. Sie geben die Versionsnummer zusammen mit dem Release-Label an. Versionsbezeichnungen haben die Form `emr-x.x.x`. z. B. `emr-7.1.0`.

Neue Amazon-EMR-Versionen werden über einen Zeitraum von mehreren Tagen in verschiedenen Regionen verfügbar gemacht, beginnend mit der ersten Region am ursprünglichen

Veröffentlichungsdatum. Die neueste Version ist in diesem Zeitraum möglicherweise nicht in Ihrer Region verfügbar.

Eine umfassende Tabelle der Anwendungsversionen in jeder Amazon-EMR-7.x-Version finden Sie unter [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#).

Themen

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Amazon EMR Version 7.1.0](#)
- [Amazon-EMR-Version 7.0.0](#)

## Anwendungsversionen in Amazon-EMR-7.x-Versionen

Eine umfassende Tabelle mit den in jeder Amazon-EMR-7.x-Version verfügbaren Anwendungsversionen finden Sie unter [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#) in Ihrem Browser.

## Amazon EMR Version 7.1.0

### 7.1.0-Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [AmazonCloudWatchAgent](#), [Delta](#), [Flink](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

## Informationen zur Anwendungsversion

	emr-7.1.0	emr-7.0.0
AWS SDK for Java	2.23.18, 1.12.656	2,20.160-amzn-0, 1,12.569
Python	3.9	3.9
Scala	2.12,17	2.12,17
AmazonCloudWatchAgent	1,300032,2	1,300031,1
Delta	3.0.0	3.0.0
Flink	1,18,1	1.18.0
Ganglia	-	-
HBase	2.4.17	2.4.17
HCatalog	3.1.3	3.1.3
Hadoop	3.3.6	3.3.6
Hive	3.1.3	3.1.3
Hudi	0.14.1-amzn-0	0.14.0-amzn-1
Hue	4.11.0	4.11.0
Iceberg	1.4.3-amzn-0	1.4.2-amzn-0
JupyterEnterpriseGateway	2.6.0	2.6.0
JupyterHub	1.5.0	1.5.0
Livy	0.8.0	0.7.1
MXNet	1.9.1	1.9.1
Mahout	-	-
Oozie	5.2.1	5.2.1

	emr-7.1.0	emr-7.0.0
Phoenix	5.1.3	5.1.3
Pig	0.17.0	0.17.0
Presto	0,284	0,283
Spark	3.5.0	3.5.0
Sqoop	1.4.7	1.4.7
TensorFlow	2.11.0	2.11.0
Tez	0,102	0.10.2
Trino (PrestoSQL)	435	426
Zeppelin	0,10,1	0.10.1
ZooKeeper	3.9.1	3.5.10

## Versionshinweise zu 7.1.0

Die folgenden Versionshinweise enthalten Informationen für Amazon EMR Version 7.1.0. Die Änderungen beziehen sich auf 7.1.0.

### Neue Features

- Anwendungs-Upgrades – Amazon EMR 7.1.0 application upgrades include Livy 0.8.0, Trino 435, and ZooKeeper 3.9.1.
- [Austausch fehlerhafter Knoten](#) — In Amazon EMR 7.1.0 und höher ist der Austausch fehlerhafter Knoten standardmäßig aktiviert, sodass Amazon EMR Ihre fehlerhaften Knoten ordnungsgemäß ersetzt. Um zu vermeiden, dass Ihre bestehenden Workflows auf Amazon EMR-Versionen 7.0.0 und niedriger beeinträchtigt werden, wird der Austausch fehlerhafter Knoten deaktiviert, wenn Sie den Kündigungsschutz in Ihrem Cluster deaktiviert haben.
- [CloudWatch Agent](#) — Konfigurieren Sie den CloudWatch Agenten so, dass er zusätzliche Systemmetriken verwendet, Anwendungsmetriken hinzufügt und das Metrikziel mit der Amazon EMR-Konfigurations-API ändert.

## Bekannte Probleme

- Python 3.11 wird von EMR Studio nicht unterstützt.

## Änderungen, Verbesserungen und behobene Probleme

- Während Amazon EMR 7.1.0 Python 3.9 standardmäßig unterstützt, unterstützen Livy 0.8.0 und Spark in Amazon EMR 7.1.0 Python 3.11.
- Diese Version behebt das Problem, dass bei Verwendung PySpark mit Python Version 3.11 jede Zeile einzeln ausgeführt werden muss.
- Zeppelin-Upgrade — Amazon EMR 7.1.0 beinhaltet ein Upgrade von Zeppelin auf das SDK for Java v2. AWS Dieses Upgrade ermöglicht es einem Zeppelin S3-Notebook, einen Anbieter von kundenspezifischen Verschlüsselungsmaterialien zu akzeptieren. Das AWS SDK for Java v2 entfernt die EncryptionMaterialsProvider Schnittstelle. Wenn Sie ein Upgrade auf Amazon EMR 7.1.0 durchführen, müssen Sie die [Keyring-Schnittstelle](#) implementieren, wenn Sie eine benutzerdefinierte Verschlüsselung verwenden möchten. [Ein Beispiel für die Implementierung der Keyring-Schnittstelle finden Sie unter .java. KmsKeyring](#)
- Wenn Sie auf Amazon EMR Version 7.1.0 aktualisieren, ändern Sie Ihren benutzerdefinierten Schlüsselanbieter für die lokale Festplattenverschlüsselung, sodass Schlüssel mithilfe des AES-Algorithmus von AES/GCM/ generiert werden. NoPadding [Wenn Sie den Algorithmus nicht aktualisieren, schlägt die Clustererstellung möglicherweise mit der folgenden Fehlermeldung fehl. Weitere Informationen Local disk encryption failed on master instance \(i-123456789\) due to internal error. zum Erstellen eines benutzerdefinierten Schlüsselanbieters finden Sie unter Erstellen eines benutzerdefinierten Schlüsselanbieters.](#)
- Amazon EMR 7.1.0 verbessert die Ausfallsicherheit eines Knotens bei geringem Festplattenspeicher, indem die Logik zur Protokollkürzung für Dateien mit offenen Datei-Handles verbessert wird.
- Diese Version verbessert die Kodierungs- und Dekodierungslogik, um das Risiko einer Datenbeschädigung und eines Knotenausfalls zu minimieren, da Amazon EMR-Daemons beim Neustart eines Knotens Dateien lesen und schreiben.
- Wenn Sie einen Cluster mit der neuesten Patch-Version von Amazon EMR 5.36 oder höher, 6.6 oder höher oder 7.0 oder höher starten, verwendet Amazon EMR die neueste Version von Amazon Linux 2023 oder Amazon Linux 2 für das standardmäßige Amazon EMR-AMI. Weitere

Informationen finden Sie unter [Verwenden des standardmäßigen Amazon-Linux-AMI für Amazon EMR](#).



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2023.3.2 240219.0	6.1.77-99.164.amzn2023	8. Mai 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), Kanada West (Calgary), AWS

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			GovCloud (US-West), AWS GovCloud (US-Ost), China (Peking), China (Ningxia)

## 7.1.0 Standard-Java-Versionen

Amazon EMR Version 7.1 und höher wird standardmäßig mit Amazon Corretto 17 (basiert auf OpenJDK) für Anwendungen geliefert, die Corretto 17 (JDK 17) unterstützen, mit Ausnahme von Apache Livy.

Die folgende Tabelle zeigt die Standard-Java-Versionen für Anwendungen in Amazon EMR 7.1.0. Wenn Sie die Standard-JVM auf Ihrem Cluster ändern möchten, folgen Sie den Anweisungen unter [Anwendungen für die Verwendung einer bestimmten Java Virtual Machine konfigurieren](#) für jede Anwendung, die auf dem Cluster ausgeführt wird. Sie können nur eine Java-Laufzeit-Version für einen Cluster verwenden. Amazon EMR unterstützt nicht die Ausführung verschiedener Knoten oder Anwendungen auf verschiedenen Laufzeitversionen auf demselben Cluster.

Anwendung	Java-/Amazon-Corretto-Version (Standard ist fett gedruckt)
Delta	17, 11, 8
Flink	17, 11, 8
Ganglia	8
HBase	11, 8
HCatalog	17, 11, 8

Anwendung	Java-/Amazon-Corretto-Version (Standard ist fett gedruckt)
Hadoop	17, 11, 8
Hive	17, 11, 8
Hudi	17, 11, 8
Iceberg	17, 11, 8
Livy	17, 11, 8
Oozie	17, 11, 8
Phoenix	8
PrestoDB	8
Spark	17, 11, 8
Spark RAPIDS	17, 11, 8
Sqoop	8
Tez	17, 11, 8
Trino	17
Zeppelin	8
Pig	8
Zookeeper	8

## 7.1.0-Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version

sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen myapp-component der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als 2.2-amzn-2 aufgeführt.

Komponente	Version	Beschreibung
adot-java-agent	1.31.0	Ein Java-Agent, der Metriken von Anwendungs-Daemons sammelt.
delta	3.0.0	Delta Lake ist ein offenes Tabellenformat für riesige analytische Datensätze
emr-amazon-cloudwatch-agent	1,300032,2-amzn-0	Eine Anwendung, die interne Metriken auf Systemebene und benutzerdefinierte Anwendungsmetriken von Amazon-EC2-Instances erfasst.
emr-ddb	5.3.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	3.10.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3,14.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-notebook-env	1.7.0	Conda Env für EMR-Notebooks, das Jupyter Enterprise Gateway enthält
emr-s3-dist-cp	2.31.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	2.10.0	EMR S3Select-Konnektor
emr-wal-cli	1.2.2	Cli wird für die emrwal list/deletion verwendet.
emrfs	2,62,0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.18.1-amzn-0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
flink-jobmanager-config	1.18.1-amzn-0	Verwaltung von Ressourcen auf EMR-Knoten für Apache JobManager Flink.
hadoop-client	3.3.6-amzn-3	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	3.3.6-amzn-3	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	3.3.6-amzn-3	HDFS-Client und -Bibliothek für die Befehlszeile

Komponente	Version	Beschreibung
hadoop-hdfs-namenode	3.3.6-amzn-3	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-hdfs-journalnode	3.3.6-amzn-3	HDFS-Service zum Verwalten des Hadoop-Dateisystem journals auf HA-Clustern.
hadoop-httfs-server	3.3.6-amzn-3	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	3.3.6-amzn-3	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	3.3.6-amzn-3	MapReduce Execution Engine-Bibliotheken zum Ausführen einer Anwendung. MapReduce
hadoop-yarn-nodemanager	3.3.6-amzn-3	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	3.3.6-amzn-3	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	3.3.6-amzn-3	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.

Komponente	Version	Beschreibung
hbase-hmaster	2.4.17-amzn-5	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	2.4.17-amzn-5	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	2.4.17-amzn-5	HBase-Befehlszeilen-Client.
hbase-rest-server	2.4.17-amzn-5	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	2.4.17-amzn-5	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hbase-operator-tools	2.4.17-amzn-5	Reparaturtool für Apache-HBase-Cluster.
hcatalog-client	3.1.3-amzn-10	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	3.1.3-amzn-10	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	3.1.3-amzn-10	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	3.1.3-amzn-10	Hive-Befehlszeilen-Client.

Komponente	Version	Beschreibung
hive-hbase	3.1.3-amzn-10	Hive-hbase client.
hive-metastore-server	3.1.3-amzn-10	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	3.1.3-amzn-10	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.14.1-amzn-0	Inkrementelles Verarbeitungs-Framework zur Versorgung der Datenpipeline mit geringer Latenz und hoher Effizienz.
hudi-presto	0.14.1-amzn-0	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hudi-trino	0.14.1-amzn-0	Bündel-Bibliothek zum Ausführen von Trino mit Hudi.
hudi-spark	0.14.1-amzn-0	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.
hue-server	4.11.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
iceberg	1.4.3-amzn-0	Apache Iceberg ist ein offenes Tabellenformat für sehr große analytische Datensätze



Komponente	Version	Beschreibung
jupyterhub	1.5.0	Multi-User-Server für Jupyter-Notebooks
livy-server	0.8.0-inkubieren	REST-Schnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP- und Reverse-Proxy-Server.
mxnet	1.9.1	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.68+	MariaDB-Datenbankserver.
nvidia-cuda	11,8,0	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.2.1	Oozie-Befehlszeilen-Client.
oozie-server	5.2.1	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	4.7.0	Open Source Computer Vision Library.
phoenix-library	5.1.3	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-connectors	5.1.3	Apache-Phoenix-Konnektoren für Spark-3

Komponente	Version	Beschreibung
phoenix-query-server	5.1.3	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0,284-amzn-0	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0,284-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0,284-amzn-0	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
trino-coordinator	435-amzn-0	Service zur Annahme von Abfragen und Verwaltung der Abfrageausführung der Trino-Worker.
trino-worker	435-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
trino-client	435-amzn-0	Trino-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Trino-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.

Komponente	Version	Beschreibung
r	4.3.2	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	2.0.0	Apache Ranger Key Management System
spark-client	3.5.0-amzn-1	Spark-Befehlszeilen-Clients.
spark-history-server	3.5.0-amzn-1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	3.5.0-amzn-1	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	3.5.0-amzn-1	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
spark-rapids	23.10.0-amzn-1	Nvidia-Spark-RAPIDS-Plugin, das Apache Spark mit GPUs beschleunigt.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	2.11.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.10.2-amzn-8	Die Tez-YARN-Anwendung und -Bibliotheken.

Komponente	Version	Beschreibung
tez-on-worker	0.10.2-amzn-8	Die Tez-YARN-Anwendung und Bibliotheken für Worker-Knoten.
zeppelin-server	0.10.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.9.1-amzn-0	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.9.1-amzn-0	ZooKeeper Befehlszeilen-Cli-ent.

## 7.1.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

### emr-7.1.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-executor	Ändern Sie die Werte in der Datei „ <code>container-executor.cfg</code> “ Datei von Hadoop YARN.

Klassifizierungen	Beschreibung
container-log4j	Ändert die Werte in der container-log4j.properties-Datei in Hadoop YARN.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.
docker-conf	Ändern Sie die Docker-bezogenen Einstellungen.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-session	Ändern Sie die Einstellungen von Flink log4j-session.properties für Kubernetes-/Yarn-Sitzungen.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.

Klassifizierungen	Beschreibung
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-env	Ändert die Werte in der HDFS-Umgebung.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der XML-Datei von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.

Klassifizierungen	Beschreibung
hive	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Hive.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.

Klassifizierungen	Beschreibung
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-java-home	Das KMS-Java-Home von Hadoop ändern
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
hudi-env	Ändern der Werte in der Hudi-Umgebung.
hudi-defaults	Ändern Sie die Werte in der hudi-defaults.conf-Datei in Hudi.
iceberg-defaults	Ändern Sie die Werte in der iceberg-defaults.conf-Datei von Iceberg.
delta-defaults	Ändern Sie die Werte in der delta-defaults.conf-Datei von Delta.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.



Klassifizierungen	Beschreibung
livy-log4j2	Ändern Sie die log4j2.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.

Klassifizierungen	Beschreibung
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-lakeformation	Ändern Sie die Werte in der lakeformation.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.
trino-log	Ändern Sie die Werte in der log.properties-Datei in Trino.
trino-config	Ändern Sie die Werte in der config.properties-Datei in Trino.
trino-password-authenticator	Ändern Sie die Werte in der Trino-Datei password-authenticator.properties.
trino-env	Ändern Sie die Werte in der trino-env.sh-Datei von Trino.
trino-node	Ändern Sie die Werte in der node.properties-Datei in Trino.

Klassifizierungen	Beschreibung
trino-connector-blackhole	Ändern Sie die Werte in der blackhole.properties-Datei in Trino.
trino-connector-cassandra	Ändern Sie die Werte in der cassandra.properties-Datei in Trino.
trino-connector-delta	Ändern Sie die Werte in der delta.properties-Datei in Trino.
trino-connector-hive	Ändern Sie die Werte in der hive.properties-Datei in Trino.
trino-exchange-manager	Ändern Sie die Werte in der exchange-manager.properties-Datei in Trino.
trino-connector-iceberg	Ändern Sie die Werte in der iceberg.properties-Datei in Trino.
trino-connector-hudi	Ändern Sie die Werte in der hudi.properties-Datei in Trino.
trino-connector-jmx	Ändern Sie die Werte in der jmx.properties-Datei in Trino.
trino-connector-kafka	Ändern Sie die Werte in der kafka.properties-Datei in Trino.
trino-connector-localfile	Ändern Sie die Werte in der localfile.properties-Datei in Trino.
trino-connector-memory	Ändern Sie die Werte in der memory.properties-Datei in Trino.
trino-connector-mongodb	Ändern Sie die Werte in der mongodb.properties-Datei in Trino.
trino-connector-mysql	Ändern Sie die Werte in der mysql.properties-Datei in Trino.

Klassifizierungen	Beschreibung
trino-connector-postgresql	Ändern Sie die Werte in der postgresql.properties-Datei in Trino.
trino-connector-raptor	Ändern Sie die Werte in der raptor.properties-Datei in Trino.
trino-connector-redis	Ändern Sie die Werte in der redis.properties-Datei in Trino.
trino-connector-redshift	Ändern Sie die Werte in der redshift.properties-Datei in Trino.
trino-connector-tpch	Ändern Sie die Werte in der tpch.properties-Datei in Trino.
trino-connector-tpcds	Ändern Sie die Werte in der tpcds.properties-Datei in Trino.
ranger-kms-dbks-site	Ändert die Werte in der dbks-site.xml-Datei von Ranger KMS.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger KMS.
ranger-kms-env	Ändert die Werte in der Ranger KMS-Umgebung.
ranger-kms-logback	Ändern Sie die Werte in der kms-logback.xml-Datei von Ranger KMS.
ranger-kms-db-ca	Ändert die Werte für die CA-Datei auf S3 für die MySQL SSL-Verbindung mit Ranger KMS.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.

Klassifizierungen	Beschreibung
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j2	Ändern Sie die Werte in der log4j2.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zeppelin-site	Ändern Sie die Konfigurationseinstellungen in zeppelin-site.xml.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.

Klassifizierungen	Beschreibung
zookeeper-logback	Ändern Sie die Werte in ZooKeeper der Datei logback.xml.
emr-metrics	Ändern Sie die emr-Metrikeinstellungen für diesen Knoten.

## 7.1.0 Änderungsprotokoll

### Änderungsprotokoll für Version 7.1.0 und Versionshinweise

Datum	Ereignis	Beschreibung
2024-05-10	Veröffentlichung von Dokumenten	Erste Veröffentlichung der Versionshinweise zu Amazon EMR 7.1.0
2024-05-08	Die Bereitstellung ist abgeschlossen	<a href="#">Amazon EMR 7.1.0 wurde vollständig in allen unterstützten Regionen bereitgestellt</a>
2024-04-23	Erstversion	Amazon EMR 7.1.0 wurde zuerst in den ersten kommerziellen Regionen eingesetzt

## Amazon-EMR-Version 7.0.0

### 7.0.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [AmazonCloudWatchAgent](#), [Delta](#), [Flink](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

#### Informationen zur Anwendungsversion

	emr-7.1.0	emr-7.0.0
AWS SDK for Java	2.23.18, 1.12.656	2,20.160-amzn-0, 1,12.569
Python	3.9	3.9
Scala	2.12,17	2.12,17
AmazonCloudWatchAgent	1,300032,2	1,300031,1
Delta	3.0.0	3.0.0
Flink	1,18,1	1.18.0
Ganglia	-	-
HBase	2.4.17	2.4.17
HCatalog	3.1.3	3.1.3
Hadoop	3.3.6	3.3.6
Hive	3.1.3	3.1.3
Hudi	0.14.1-amzn-0	0.14.0-amzn-1



	emr-7.1.0	emr-7.0.0
Hue	4.11.0	4.11.0
Iceberg	1.4.3-amzn-0	1.4.2-amzn-0
JupyterEnterpriseGateway	2.6.0	2.6.0
JupyterHub	1.5.0	1.5.0
Livy	0.8.0	0.7.1
MXNet	1.9.1	1.9.1
Mahout	-	-
Oozie	5.2.1	5.2.1
Phoenix	5.1.3	5.1.3
Pig	0.17.0	0.17.0
Presto	0,284	0,283
Spark	3.5.0	3.5.0
Sqoop	1.4.7	1.4.7
TensorFlow	2.11.0	2.11.0
Tez	0,102	0.10.2
Trino (PrestoSQL)	435	426
Zeppelin	0,10,1	0.10.1
ZooKeeper	3.9.1	3.5.10

## 7.0.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 7.0.0. Änderungen beziehen sich auf Version 6.15.0.

### Neue Features

- **Anwendungs-Upgrades** – Amazon EMR 7.0.0 application upgrades include Python 3.9, Spark 3.5, Flink 1.18, and Delta 3.0. Diese Version bietet auch Unterstützung für die Anwendung [CloudWatch Amazon-Vertreter](#) und entfernt die Unterstützung für Ganglia.
- **Amazon Corretto 17** – Amazon-EMR-Versionen 7.0 und höher werden standardmäßig mit Amazon Corretto 17 (basiert auf OpenJDK) für Anwendungen ausgeliefert, die Corretto 17 (JDK 17) unterstützen, mit Ausnahme von Apache Livy. Weitere Informationen zu den unterstützten JDK-Versionen für Anwendungen in diesem Release finden Sie unter [7.0.0 Standard-Java-Versionen](#).
- **Amazon Linux 2023** – Mit der Version 7.0 werden Amazon-EMR-Cluster jetzt standardmäßig in AL2023 ausgeführt. Informationen dazu, wie sich dies auf die Standard-AMI-Version auswirkt, finden Sie unter [Überlegungen zum Software-Update](#) im Leitfaden zu Amazon EMR. Beachten Sie auch, dass AL2023 Python 2.7 entfernt hat, sodass alle Komponenten, die Python benötigen, jetzt mit Python 3 geschrieben werden sollten.
- **S3 in Outposts mit s3a** – Amazon EMR unterstützt jetzt Amazon S3 in Outposts-Buckets mit dem s3a-Dateisystem. Weitere Informationen zu S3 in Outposts finden Sie unter [Was ist S3 in Outposts?](#) im Benutzerhandbuch für Amazon Simple Storage Service.

### Bekannte Probleme

- Beachten Sie, dass Sie nicht mehr als eine Zeile gleichzeitig ausführen können, wenn Sie Python-Versionen 3.10 oder höher verwenden PySpark . Sie müssen jede Zeile einzeln ausführen.

### Änderungen, Verbesserungen und behobene Probleme

- Um den Status aller Instances in einem Cluster aufrechtzuerhalten, ist Amazon EMR in Apache YARN, Apache HDFS und Kerberos integriert. Mit 7.0 haben wir diese Integrationen verbessert, um den Status von Instances, die aufgrund von Skalierung oder anderen Vorgängen beendet wurden, zuverlässig zu entfernen. Dies ist besonders wichtig für langlebige Cluster, bei denen die verwaltete Skalierung aktiviert ist, da sie häufig skalieren und im Laufe ihrer Lebensdauer Tausende beendete Instances ansammeln.

- In dieser Version wurde die Kerberos-Konfiguration dahingehend verbessert, dass nur AES-basierte Verschlüsselungen unterstützt werden. Kerberos KDC mit Verschlüsselungen, die nicht auf AES basieren, werden von EMR-Clustern, die in Amazon-EMR-Versionen 7.0.0 und höher ausgeführt werden, nicht mehr unterstützt. AES-basierte Verschlüsselungen bieten die stärkste Sicherheit für Ihre Cluster.
- Im Rahmen der AWS SDK 2.x-Migration enthält Amazon EMR 7.0 aus Kompatibilitätsgründen ein Update für den Spark Kinesis-Konnektor. Dieses Update ist in der Community-Version von Apache Spark nicht verfügbar. Wenn Sie den Spark-Kinesis-Konnektor aus einer Amazon-EMR-Version unter 7.0 verwenden, müssen Sie Ihre Anwendungscode für die Ausführung in SDK 2.x migrieren, bevor Sie Ihre Workloads zu Amazon EMR 7.0 migrieren können. Weitere Informationen finden Sie unter [Migration des Spark-Kinesis-Konnektors zu SDK 2.x für Amazon EMR 7.0](#).
- Wenn Sie einen Cluster mit der neuesten Patch-Version von Amazon EMR 5.36 oder höher, 6.6 oder höher oder 7.0 oder höher starten, verwendet Amazon EMR die neueste Version von Amazon Linux 2023 oder Amazon Linux 2 für das standardmäßige Amazon EMR-AMI. Weitere Informationen finden Sie unter [Verwenden des standardmäßigen Amazon-Linux-AMI für Amazon EMR](#).

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2023.3.2 240304.1	6.1.79-99.164.amzn2023	12. März 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Paz

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			ifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), Kanada West (Calgary), AWS GovCloud (US-West), AWS GovCloud (US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2023.3.240219.0	6.1.77-99.164.amzn2023	1. März 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), Kanada West (Calgary), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			AWS GovCloud (US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2023.3.240205,1	6.1.75-99.163.amzn2023	19. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), Kanada West (Calgary), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			AWS GovCloud (US-Ost), China (Peking), China (Ningxia)



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2023.3.240122,0	6.1.72-96.amzn2023	5. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), Kanada West (Calgary), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			AWS GovCloud (US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2023.3.240108,1	6.1.72-96.amzn2023	24. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), Kanada West (Calgary), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			AWS GovCloud (US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2023.3.231211,4	6.1.66-91.160.amzn2023	19. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

## 7.0.0 Standard-Java-Versionen

Amazon-EMR-Version 7.0 und höher wird standardmäßig mit Amazon Corretto 17 (basiert auf OpenJDK) für Anwendungen ausgeliefert, die Corretto 17 (JDK 17) unterstützen, mit Ausnahme von Apache Livy.

Die folgende Tabelle zeigt die Standard-Java-Versionen für Anwendungen in Amazon EMR 7.0.0. Wenn Sie die Standard-JVM auf Ihrem Cluster ändern möchten, folgen Sie den Anweisungen unter [Anwendungen für die Verwendung einer bestimmten Java Virtual Machine konfigurieren](#) für jede Anwendung, die auf dem Cluster ausgeführt wird. Sie können nur eine Java-Laufzeit-Version für einen Cluster verwenden. Amazon EMR unterstützt nicht die Ausführung verschiedener Knoten oder Anwendungen auf verschiedenen Laufzeitversionen auf demselben Cluster.

Anwendung	Java-/Amazon-Corretto-Version (Standard ist fett gedruckt)
Delta	17, 11, 8
Flink	11, 8
Ganglia	8
HBase	11, 8
HCatalog	17, 11, 8
Hadoop	17, 11, 8

Anwendung	Java-/Amazon-Corretto-Version (Standard ist fett gedruckt)
Hive	17, 11, 8
Hudi	17, 11, 8
Iceberg	17, 11, 8
Livy	17, 11, 8
Oozie	17, 11, 8
Phoenix	8
PrestoDB	8
Spark	17, 11, 8
Spark RAPIDS	17, 11, 8
Sqoop	8
Tez	17, 11, 8
Trino	17
Zeppelin	8
Pig	8
Zookeeper	8

## 7.0.0 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version

sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen myapp-component der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als 2.2-amzn-2 aufgeführt.

Komponente	Version	Beschreibung
adot-java-agent	1,31,0	Ein Java-Agent, der Metriken von Anwendungs-Daemons sammelt.
delta	3.0.0	Delta Lake ist ein offenes Tabellenformat für riesige analytische Datensätze
delta-standalone-connectors	0.6.0	Delta-Konnektoren bieten unterschiedliche Laufzeiten, um Delta Lake mit Engines wie Flink, Hive und Presto zu integrieren.
emr-amazon-cloudwatch-agent	1.300031.1-amzn-0	Eine Anwendung, die interne Metriken auf Systemebene und benutzerdefinierte Anwendungsmetriken von Amazon-EC2-Instances erfasst.
emr-ddb	5.2.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.



Komponente	Version	Beschreibung
emr-goodies	3.9.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.13,0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-notebook-env	1.7.0	Conda Env für EMR-Notebooks, das Jupyter Enterprise Gateway enthält
emr-s3-dist-cp	2.30.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	2.9.0	EMR S3Select-Konnektor
emr-wal-cli	1.2.1	Cli wird für die emrwal list/deletion verwendet.
emrfs	2,61,0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.18.0-amzn-0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
flink-jobmanager-config	1.18.0-amzn-0	Verwaltung von Ressourcen auf EMR-Knoten für Apache JobManager Flink.
hadoop-client	3.3.6-amzn-2	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".

Komponente	Version	Beschreibung
hadoop-hdfs-datanode	3.3.6-amzn-2	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	3.3.6-amzn-2	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	3.3.6-amzn-2	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-hdfs-journalnode	3.3.6-amzn-2	HDFS-Service zum Verwalten des Hadoop-Dateisystem journals auf HA-Clustern.
hadoop-httfs-server	3.3.6-amzn-2	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	3.3.6-amzn-2	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	3.3.6-amzn-2	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	3.3.6-amzn-2	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	3.3.6-amzn-2	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.

Komponente	Version	Beschreibung
hadoop-yarn-timeline-server	3.3.6-amzn-2	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	2.4.17-amzn-4	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	2.4.17-amzn-4	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	2.4.17-amzn-4	HBase-Befehlszeilen-Client.
hbase-rest-server	2.4.17-amzn-4	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	2.4.17-amzn-4	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hbase-operator-tools	2.4.17-amzn-4	Reparaturtool für Apache-HBase-Cluster.
hcatalog-client	3.1.3-amzn-9	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	3.1.3-amzn-9	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).

Komponente	Version	Beschreibung
hcatalog-webhcat-server	3.1.3-amzn-9	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	3.1.3-amzn-9	Hive-Befehlszeilen-Client.
hive-hbase	3.1.3-amzn-9	Hive-hbase client.
hive-metastore-server	3.1.3-amzn-9	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	3.1.3-amzn-9	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.14.0-amzn-1	Inkrementelles Verarbeitungs-Framework zur Versorgung der Datenpipeline mit geringer Latenz und hoher Effizienz.
hudi-presto	0.14.0-amzn-1	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hudi-trino	0.14.0-amzn-1	Bündel-Bibliothek zum Ausführen von Trino mit Hudi.
hudi-spark	0.14.0-amzn-1	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.
hue-server	4.11.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.

Komponente	Version	Beschreibung
iceberg	1.4.2-amzn-0	Apache Iceberg ist ein offenes Tabellenformat für sehr große analytische Datensätze
jupyterhub	1.5.0	Multi-User-Server für Jupyter-Notebooks
livy-server	0.7.1-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP- und Reverse-Proxy-Server.
mxnet	1.9.1	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.68+	MariaDB-Datenbankserver.
nvidia-cuda	11,8,0	Nvidia-Treiber und Cuda-Toolkit
oozie-client	5.2.1	Oozie-Befehlszeilen-Client.
oozie-server	5.2.1	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	4.7.0	Open Source Computer Vision Library.
phoenix-library	5.1.3	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-connectors	5.1.3	Apache-Phoenix-Konnektoren für Spark-3

Komponente	Version	Beschreibung
phoenix-query-server	5.1.3	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.283-amzn-1	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.283-amzn-1	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.283-amzn-1	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
trino-coordinator	426-amzn-1	Service zur Annahme von Abfragen und Verwaltung der Abfrageausführung der Trino-Worker.
trino-worker	426-amzn-1	Service für das Ausführen von Teilen einer Abfrage.
trino-client	426-amzn-1	Trino-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Trino-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.

Komponente	Version	Beschreibung
r	4.1.3	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	2.0.0	Apache Ranger Key Management System
spark-client	3.5.0-amzn-0	Spark-Befehlszeilen-Clients.
spark-history-server	3.5.0-amzn-0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	3.5.0-amzn-0	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	3.5.0-amzn-0	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
spark-rapids	23.10.0-amzn-0	Nvidia-Spark-RAPIDS-Plugin, das Apache Spark mit GPUs beschleunigt.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	2.11.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.10.2-amzn-7	Die Tez-YARN-Anwendung und -Bibliotheken.

Komponente	Version	Beschreibung
tez-on-worker	0.10.2-amzn-7	Die Tez-YARN-Anwendung und Bibliotheken für Worker-Knoten.
webserver	2.4.58	Apache HTTP-Server.
zeppelin-server	0.10.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.5.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.5.10	ZooKeeper Befehlszeilen-Cli-ent.

## 7.0.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

### emr-7.0.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-executor	Ändern Sie die Werte in der Datei „ <code>container-executor.cfg</code> “ Datei von Hadoop YARN.



Klassifizierungen	Beschreibung
container-log4j	Ändert die Werte in der container-log4j.properties-Datei in Hadoop YARN.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.
docker-conf	Ändern Sie die Docker-bezogenen Einstellungen.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-session	Ändern Sie die Einstellungen von Flink log4j-session.properties für Kubernetes-/Yarn-Sitzungen.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.

Klassifizierungen	Beschreibung
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-env	Ändert die Werte in der HDFS-Umgebung.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der proto-hive-site.xml-Datei von HCatalog.
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.

Klassifizierungen	Beschreibung
hive	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Hive.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.

Klassifizierungen	Beschreibung
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-java-home	Das KMS-Java-Home von Hadoop ändern
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
hudi-env	Ändern der Werte in der Hudi-Umgebung.
hudi-defaults	Ändern Sie die Werte in der hudi-defaults.conf-Datei in Hudi.
iceberg-defaults	Ändern Sie die Werte in der iceberg-defaults.conf-Datei von Iceberg.
delta-defaults	Ändern Sie die Werte in der delta-defaults.conf-Datei von Delta.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.

Klassifizierungen	Beschreibung
livy-log4j2	Ändern Sie die log4j2.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.

Klassifizierungen	Beschreibung
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-lakeformation	Ändern Sie die Werte in der lakeformation.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.
trino-log	Ändern Sie die Werte in der log.properties-Datei in Trino.
trino-config	Ändern Sie die Werte in der config.properties-Datei in Trino.
trino-password-authenticator	Ändern Sie die Werte in der Trino-Datei password-authenticator.properties.
trino-env	Ändern Sie die Werte in der trino-env.sh-Datei von Trino.
trino-node	Ändern Sie die Werte in der node.properties-Datei in Trino.

Klassifizierungen	Beschreibung
trino-connector-blackhole	Ändern Sie die Werte in der blackhole.properties-Datei in Trino.
trino-connector-cassandra	Ändern Sie die Werte in der cassandra.properties-Datei in Trino.
trino-connector-delta	Ändern Sie die Werte in der delta.properties-Datei in Trino.
trino-connector-hive	Ändern Sie die Werte in der hive.properties-Datei in Trino.
trino-exchange-manager	Ändern Sie die Werte in der exchange-manager.properties-Datei in Trino.
trino-connector-iceberg	Ändern Sie die Werte in der iceberg.properties-Datei in Trino.
trino-connector-hudi	Ändern Sie die Werte in der hudi.properties-Datei in Trino.
trino-connector-jmx	Ändern Sie die Werte in der jmx.properties-Datei in Trino.
trino-connector-kafka	Ändern Sie die Werte in der kafka.properties-Datei in Trino.
trino-connector-localfile	Ändern Sie die Werte in der localfile.properties-Datei in Trino.
trino-connector-memory	Ändern Sie die Werte in der memory.properties-Datei in Trino.
trino-connector-mongodb	Ändern Sie die Werte in der mongodb.properties-Datei in Trino.
trino-connector-mysql	Ändern Sie die Werte in der mysql.properties-Datei in Trino.



Klassifizierungen	Beschreibung
trino-connector-postgresql	Ändern Sie die Werte in der postgresql.properties-Datei in Trino.
trino-connector-raptor	Ändern Sie die Werte in der raptor.properties-Datei in Trino.
trino-connector-redis	Ändern Sie die Werte in der redis.properties-Datei in Trino.
trino-connector-redshift	Ändern Sie die Werte in der redshift.properties-Datei in Trino.
trino-connector-tpch	Ändern Sie die Werte in der tpch.properties-Datei in Trino.
trino-connector-tpcds	Ändern Sie die Werte in der tpcds.properties-Datei in Trino.
ranger-kms-dbks-site	Ändert die Werte in der dbks-site.xml-Datei von Ranger KMS.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger KMS.
ranger-kms-env	Ändert die Werte in der Ranger KMS-Umgebung.
ranger-kms-logback	Ändern Sie die Werte in der kms-logback.xml-Datei von Ranger KMS.
ranger-kms-db-ca	Ändert die Werte für die CA-Datei auf S3 für die MySQL SSL-Verbindung mit Ranger KMS.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.

Klassifizierungen	Beschreibung
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j2	Ändern Sie die Werte in der log4j2.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zeppelin-site	Ändern Sie die Konfigurationseinstellungen in zeppelin-site.xml.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.

Klassifizierungen	Beschreibung
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## 7.0.0 Änderungsprotokoll

### Änderungsprotokoll für Version 7.0.0 und Versionshinweise

Datum	Ereignis	Beschreibung
2023-12-29	Die Bereitstellung ist abgeschlossen	<a href="#">Amazon EMR 7.0.0 wurde vollständig in allen unterstützten Regionen bereitgestellt</a>
28.12.2023-12	Veröffentlichung von Dokumenten	Erste Veröffentlichung der Versionshinweise zu Amazon EMR 7.0.0
2023-12-19	Erstversion	Amazon EMR 7.0.0 wurde zuerst in ersten kommerziellen Regionen bereitgestellt

## Amazon-EMR-6.x-Versionen

Dieser Abschnitt enthält Anwendungsversionen, Versionshinweise, Komponentenversionen und Konfigurationsklassifizierungen, die in jeder Amazon-EMR-6.x-Version verfügbar sind.

Wenn Sie einen Cluster starten, können Sie aus verschiedenen Versionen von Amazon EMR wählen. Auf diese Weise können Sie Anwendungsversionen testen und verwenden, die zu den Kompatibilitätsanforderungen passen. Sie geben die Versionsnummer zusammen mit dem Release-Label an. Versionsbezeichnungen haben die Form `emr-x.x.x`. z. B. `emr-7.1.0`.

Neue Amazon-EMR-Versionen werden über einen Zeitraum von mehreren Tagen in verschiedenen Regionen verfügbar gemacht, beginnend mit der ersten Region am ursprünglichen Veröffentlichungsdatum. Die neueste Version ist in diesem Zeitraum möglicherweise nicht in Ihrer Region verfügbar.

Eine umfassende Tabelle der Anwendungsversionen in jeder Amazon-EMR-6.x-Version finden Sie unter [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#).

## Themen

- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Amazon-EMR-Version 6.15.0](#)
- [Amazon-EMR-Version 6.14.0](#)
- [Amazon-EMR-Version 6.13.0](#)
- [Amazon-EMR-Version 6.12.0](#)
- [Amazon-EMR-Version 6.11.1](#)
- [Amazon-EMR-Version 6.11.0](#)
- [Amazon-EMR-Version 6.10.1](#)
- [Amazon-EMR-Version 6.10.0](#)
- [Amazon-EMR-Version 6.9.1](#)
- [Amazon-EMR-Version 6.9.0](#)
- [Amazon-EMR-Version 6.8.1](#)
- [Amazon-EMR-Version 6.8.0](#)
- [Amazon-EMR-Version 6.7.0](#)
- [Amazon-EMR-Version 6.6.0](#)
- [Amazon-EMR-Version 6.5.0](#)
- [Amazon-EMR-Version 6.4.0](#)
- [Amazon-EMR-Version 6.3.1](#)
- [Amazon-EMR-Version 6.3.0](#)
- [Amazon-EMR-Version 6.2.1](#)
- [Amazon-EMR-Version 6.2.0](#)
- [Amazon-EMR-Version 6.1.1](#)
- [Amazon-EMR-Version 6.1.0](#)
- [Amazon-EMR-Version 6.0.1](#)
- [Amazon-EMR-Version 6.0.0](#)

## Anwendungsversionen in Amazon-EMR-6.x-Versionen

Eine umfassende Tabelle mit den in jeder Amazon-EMR-6.x-Version verfügbaren Anwendungsversionen finden Sie unter [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#) in Ihrem Browser.

### Amazon-EMR-Version 6.15.0

#### 6.15.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Delta](#), [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

#### Informationen zur Anwendungsversion

	emr-6.15.0	emr-6.14.0	emr-6.13.0	emr-6.12.0
AWS SDK for Java	2,20.160-amzn-0, 1,12.569	1.12.543	1,12,513	1.12,490
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2.12,17	2.12,15	2.12,15	2.12,15

	emr-6.15.0	emr-6.14.0	emr-6.13.0	emr-6.12.0
AmazonCloudWatchAgent	-	-	-	-
Delta	2.4.0	2.4.0	2.4.0	2.4.0
Flink	1.17.1	1.17.1	1.17.0	1.17.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4.17	2.4.17	2.4.17	2.4.17
HCatalog	3.1.3	3.1.3	3.1.3	3.1.3
Hadoop	3.3.6	3.3.3	3.3.3	3.3.3
Hive	3.1.3	3.1.3	3.1.3	3.1.3
Hudi	0.14.0-amzn-0	0.13.1-amzn-2	0.13.1-amzn-1	0.13.1-amzn-0
Hue	4.11.0	4.11.0	4.11.0	4.11.0
Iceberg	1.4.0-amzn-0	1.3.1-amzn-0	1.3.0-amzn-1	1.3.0-amzn-0
JupyterEnterpriseGateway	2.6.0	2.6.0	2.6.0	2.6.0
JupyterHub	1.5.0	1.5.0	1.5.0	1.4.1
Livy	0.7.1	0.7.1	0.7.1	0.7.1
MXNet	1.9.1	1.9.1	1.9.1	1.9.1
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	5.1.3	5.1.3	5.1.3	5.1.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0

	emr-6.15.0	emr-6.14.0	emr-6.13.0	emr-6.12.0
Presto	0,283	0,281	0,281	0,281
Spark	3.4.1	3.4.1	3.4.1	3.4.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.11.0	2.11.0	2.11.0	2.11.0
Tez	0,102	0.10.2	0.10.2	0.10.2
Trino (PrestoSQL)	426	422	414	414
Zeppelin	0,10,1	0.10.1	0.10.1	0.10.1
ZooKeeper	3.5.10	3.5.10	3.5.10	3.5.10

## 6.15.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 6.15.0. Änderungen beziehen sich auf Version 6.14.0. Informationen zum Zeitplan für die Veröffentlichung finden Sie unter [6.15.0 Änderungsprotokoll](#).

### Neue Features

- Anwendungs-Upgrades – Amazon EMR 6.15.0 application upgrades include Apache Hadoop 3.3.6, Apache Hudi 0.14.0-amzn-0, Iceberg 1.4.0-amzn-0, and Trino 426.
- [Schnellere Starts für EMR-Cluster, die in EC2 ausgeführt werden](#) – Ein Amazon EMR kann jetzt bis zu 35 % schneller in einem EC2-Cluster gestartet werden. Mit dieser Verbesserung können die meisten Kunden ihre Cluster in 5 Minuten oder weniger starten.
- [CodeWhisperer für EMR Studio](#) — Sie können Amazon jetzt CodeWhisperer mit Amazon EMR Studio verwenden, um Empfehlungen in Echtzeit zu erhalten, während Sie Code einschreiben. JupyterLab CodeWhisperer kann Ihre Kommentare vervollständigen, einzelne Codezeilen fertigstellen, line-by-line Empfehlungen aussprechen und vollständig formatierte Funktionen generieren.

- [Schnellere Auftrag-Neustartzeiten mit Flink](#) – Mit Amazon EMR 6.15.0 und höher sind mehrere neue Mechanismen für Apache Flink verfügbar, um die Auftrag-Neustartzeit bei Aufgabenwiederherstellungs- oder -skalierungsvorgängen zu verkürzen. Dadurch wird die Geschwindigkeit der Wiederherstellung und des Neustarts von Ausführungsdiagrammen optimiert, um die Auftragsstabilität zu verbessern.
- Detaillierte [Zugriffskontrolle auf Tabellenebene für Open-Table-Formate](#) — Wenn Sie mit Amazon EMR 6.15.0 und höher Spark-Jobs auf Amazon EMR auf EC2-Clustern ausführen, die auf Daten im AWS Glue-Datenkatalog zugreifen, können Sie damit Berechtigungen auf Tabellen-, Zeilen-, Spalten- und Zellenebene auf Hudi-, Iceberg- oder Delta Lake-basierte Tabellen anwenden. AWS Lake Formation
- Hadoop-Upgrade – Amazon EMR 6.15.0 beinhaltet ein Upgrade von Apache Hadoop auf Version 3.3.6. Hadoop 3.3.6 war die neueste Version zum Zeitpunkt der Bereitstellung von Amazon EMR 6.15, die im Juni 2023 von Apache veröffentlicht wurde. Frühere Versionen von Amazon EMR (6.9.0 bis 6.14.x) verwendeten Hadoop 3.3.3.

Das Upgrade umfasst Hunderte Verbesserungen und Fehlerbehebungen sowie Features wie rekonfigurierbare Datenknotenparameter, die DFSAdmin-Option zum Initiieren von Massenrekonfigurationsvorgängen für alle Live-Datenknoten und eine vektorbasierte API, die es suchlastigen Lesern ermöglicht, mehrere Lesebereiche anzugeben. Hadoop 3.3.6 bietet außerdem Unterstützung für HDFS-APIs und Semantik für sein Write Ahead Log (WAL), sodass HBase in anderen Speichersystemimplementierungen ausgeführt werden kann. Weitere Informationen finden Sie in den Änderungsprotokollen für Versionen [3.3.4](#), [3.3.5](#) und [3.3.6](#) in der Apache-Hadoop-Dokumentation.

- Support für AWS SDK for Java, Version 2 — Amazon EMR 6.15.0-Anwendungen können AWS SDK for Java Java-Versionen [1.12.569](#) oder [2.20.160](#) verwenden, wenn die Anwendung v2 unterstützt. Das AWS SDK for Java 2.x ist eine umfassende Neufassung der Codebasis von Version 1.x. Sie basiert auf Java 8+ und fügt mehrere häufig angeforderte Funktionen hinzu. Dies umfasst die Unterstützung von blockierungsfreiem E/A und die Möglichkeit, zur Laufzeit eine andere HTTP-Implementierung einzubinden. Weitere Informationen, einschließlich eines Migrationshandbuchs von SDK für Java v1 zu v2, finden Sie im Handbuch [AWS SDK für Java, Version 2](#).

## Änderungen, Verbesserungen und behobene Probleme

- Um Ihre EMR-Cluster mit hoher Verfügbarkeit zu verbessern, ermöglicht diese Version Konnektivität zu Amazon-EMR-Daemons auf lokalen Hosts, die IPv6-Endpunkte verwenden.



- Diese Version aktiviert TLS 1.2 für die Kommunikation mit ZooKeeper Provisioned auf allen primären Knoten Ihres Hochverfügbarkeitsclusters.
- Diese Version verbessert die Verwaltung von ZooKeeper Transaktionsprotokolldateien, die auf primären Knoten verwaltet werden, um Szenarien zu minimieren, in denen die Protokolldateien über die Grenzen hinaus wachsen und den Clusterbetrieb unterbrechen.
- Diese Version verbessert die Widerstandsfähigkeit der knoteninternen Kommunikation von EMR-Clustern mit hoher Verfügbarkeit. Durch diese Verbesserung wird die Wahrscheinlichkeit, dass Bootstrap-Aktionen oder Cluster-Startfehler auftreten, verringert.
- Tez in Amazon EMR 6.15.0 führt Konfigurationen ein, die Sie angeben können, um die Eingabe-Splits in einem gruppierten Tez-Split asynchron zu öffnen. Dies führt zu einer schnelleren Leistung von Leseabfragen, wenn ein einzelner gruppierter Tez-Split eine große Anzahl von Eingabe-Splits enthält. Weitere Informationen finden Sie unter [Asynchrone Öffnung von Tez-Splits](#).
- Wenn Sie einen Cluster mit der neuesten Patch-Version von Amazon EMR 5.36 oder höher, 6.6 oder höher oder 7.0 oder höher starten, verwendet Amazon EMR die neueste Version von Amazon Linux 2023 oder Amazon Linux 2 für das standardmäßige Amazon EMR-AMI. Weitere Informationen finden Sie unter [Verwenden des standardmäßigen Amazon-Linux-AMI für Amazon EMR](#).

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 223.0	4,14.336	8. März 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost), China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024131.0	4,14.336	14. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024124.0	4,14.336	7. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsReleaseLabel (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024109.0	4,14.334	24. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 218.0	4,14.330	2. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsReleaseLabel (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202306.0	4,14.330	22. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 116.0	4,14.328	11. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023101.0	4,14.327	13. November 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

## 6.15.0 Standard-Java-Versionen

Amazon-EMR-Versionen 6.12.0 und höher unterstützen standardmäßig alle Anwendungen mit Amazon Corretto 8, mit Ausnahme von Trino. Für Trino unterstützt Amazon EMR Amazon Corretto 17 ab Amazon-EMR-Version 6.9.0 standardmäßig. Amazon EMR unterstützt auch einige Anwendungen mit Amazon Corretto 11 und 17. Diese Anwendungen werden in der folgenden Tabelle aufgelistet. Wenn Sie die Standard-JVM auf Ihrem Cluster ändern möchten, folgen Sie den Anweisungen unter [Anwendungen für die Verwendung einer bestimmten Java Virtual Machine konfigurieren](#) für jede Anwendung, die auf dem Cluster ausgeführt wird. Sie können nur eine Java-Laufzeit-Version für einen Cluster verwenden. Amazon EMR unterstützt nicht die Ausführung verschiedener Knoten oder Anwendungen auf verschiedenen Laufzeitversionen auf demselben Cluster.

Amazon EMR unterstützt zwar sowohl Amazon Corretto 11 als auch 17 auf Apache Spark, Apache Hadoop und Apache Hive, bei einigen Workloads kann es jedoch zu Leistungseinbußen kommen, wenn Sie diese Versionen von Corretto verwenden. Wir empfehlen Ihnen, Ihre Workloads zu testen, bevor Sie die Standardeinstellungen ändern.

Die folgende Tabelle zeigt die Standard-Java-Versionen für Anwendungen in Amazon EMR 6.15.0:

Anwendung	Java-/Amazon-Corretto-Version (Standard ist fett gedruckt)
Delta	<b>17</b> , 11, 8
Flink	11, <b>8</b>
Ganglia	8



Anwendung	Java-/Amazon-Corretto-Version (Standard ist fett gedruckt)
HBase	11, 8
HCatalog	17, 11, 8
Hadoop	17, 11, 8
Hive	17, 11, 8
Hudi	17, 11, 8
Iceberg	17, 11, 8
Livy	17, 11, 8
Oozie	17, 11, 8
Phoenix	8
PrestoDB	8
Spark	17, 11, 8
Spark RAPIDS	17, 11, 8
Sqoop	8
Tez	17, 11, 8
Trino	17
Zeppelin	8
Pig	8
Zookeeper	8

## 6.15.0 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.4.2	Amazon SageMaker Spark-SDK
<code>delta</code>	2.4.0	Delta Lake ist ein offenes Tabellenformat für riesige analytische Datensätze
<code>delta-standalone-connectors</code>	0.6.0	Delta-Konnektoren bieten unterschiedliche Laufzeiten, um Delta Lake mit Engines wie Flink, Hive und Presto zu integrieren.
<code>emr-ddb</code>	5.2.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	3.8.0	Praktische Bibliotheken für das Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-kinesis	3.12.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-notebook-env	1.7.0	Conda Env für EMR-Notebooks, das Jupyter Enterprise Gateway enthält
emr-s3-dist-cp	2.29.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	2.8.0	EMR S3Select-Konnektor
emr-wal-cli	1.2.0	Cli wird für die emrwal list/deletion verwendet.
emrfs	2.60.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.17.1-amzn-1	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
flink-jobmanager-config	1.17.1-amzn-1	Verwaltung von Ressourcen auf EMR-Knoten für Apache JobManager Flink.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.

Komponente	Version	Beschreibung
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	3.3.6-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	3.3.6-amzn-1	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	3.3.6-amzn-1	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	3.3.6-amzn-1	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-hdfs-journalnode	3.3.6-amzn-1	HDFS-Service zum Verwalten des Hadoop-Dateisystem journals auf HA-Clustern.
hadoop-httpfs-server	3.3.6-amzn-1	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	3.3.6-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider

Komponente	Version	Beschreibung
hadoop-mapred	3.3.6-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	3.3.6-amzn-1	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	3.3.6-amzn-1	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	3.3.6-amzn-1	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	2.4.17-amzn-3	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	2.4.17-amzn-3	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	2.4.17-amzn-3	HBase-Befehlszeilen-Client.
hbase-rest-server	2.4.17-amzn-3	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	2.4.17-amzn-3	Service, der einen Thrift-Endpunkt für HBase bereitstellt.

Komponente	Version	Beschreibung
hbase-operator-tools	2.4.17-amzn-3	Reparaturtool für Apache-HBase-Cluster.
hcatalog-client	3.1.3-amzn-8	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	3.1.3-amzn-8	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	3.1.3-amzn-8	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	3.1.3-amzn-8	Hive-Befehlszeilen-Client.
hive-hbase	3.1.3-amzn-8	Hive-hbase client.
hive-metastore-server	3.1.3-amzn-8	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	3.1.3-amzn-8	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.14.0-amzn-0	Inkrementelles Verarbeitungs-Framework zur Versorgung der Datenpipeline mit geringer Latenz und hoher Effizienz.

Komponente	Version	Beschreibung
hudi-presto	0.14.0-amzn-0	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hudi-trino	0.14.0-amzn-0	Bündel-Bibliothek zum Ausführen von Trino mit Hudi.
hudi-spark	0.14.0-amzn-0	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.
hue-server	4.11.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
iceberg	1.4.0-amzn-0	Apache Iceberg ist ein offenes Tabellenformat für sehr große analytische Datensätze
jupyterhub	1.5.0	Multi-User-Server für Jupyter-Notebooks
livy-server	0.7.1-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP- und Reverse-Proxy-Server.
mxnet	1.9.1	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.68+	MariaDB-Datenbankserver.
nvidia-cuda	11,8,0	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.2.1	Oozie-Befehlszeilen-Client.

Komponente	Version	Beschreibung
oozie-server	5.2.1	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	4.7.0	Open Source Computer Vision Library.
phoenix-library	5.1.3	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-connectors	5.1.3	Apache-Phoenix-Konnektoren für Spark-3
phoenix-query-server	5.1.3	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.283-amzn-0	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.283-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.283-amzn-0	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
trino-coordinator	426-amzn-0	Service zur Annahme von Abfragen und Verwaltung der Abfrageausführung der Trino-Worker.



Komponente	Version	Beschreibung
trino-worker	426-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
trino-client	426-amzn-0	Trino-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Trino-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	4.0.2	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	2.0.0	Apache Ranger Key Management System
spark-client	3.4.1-amzn-2	Spark-Befehlszeilen-Clients.
spark-history-server	3.4.1-amzn-2	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	3.4.1-amzn-2	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	3.4.1-amzn-2	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
spark-rapids	23.08.1-amzn-0	Nvidia-Spark-RAPIDS-Plugin, das Apache Spark mit GPUs beschleunigt.

Komponente	Version	Beschreibung
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	2.11.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.10.2-amzn-6	Die Tez-YARN-Anwendung und -Bibliotheken.
tez-on-worker	0.10.2-amzn-6	Die Tez-YARN-Anwendung und Bibliotheken für Worker-Knoten.
webserver	2.4.41+	Apache HTTP-Server.
zeppelin-server	0.10.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.5.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.5.10	ZooKeeper Befehlszeilen-Client.

## 6.15.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Aktionen zur Neukonfiguration treten auf, wenn Sie eine Konfiguration für Instance-Gruppen in einem laufenden Cluster angeben. Amazon EMR initiiert nur Rekonfigurationsaktionen für die Klassifizierungen, die Sie ändern. Weitere Informationen finden Sie unter [Eine Instance-Gruppe in einem laufenden Cluster neu konfigurieren](#).

### emr-6.15.0 Klassifizierungen

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.	Restarts the Resource Manager service.
container-executor	Ändern Sie die Werte in der Datei „ <code>container-executor.cfg</code> “ Datei von Hadoop YARN.	Not available.
container-log4j	Ändert die Werte in der <code>container-log4j.properties</code> -Datei in Hadoop YARN.	Not available.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
		Httpfs, and MapReduce-HistoryServer.
docker-conf	Ändern Sie die Docker-bezogenen Einstellungen.	Not available.
emrfs-site	Ändert die EMRFS-Einstellungen.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegistratorserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.	Restarts Flink history server.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.	Restarts Flink history server.
flink-log4j-session	Ändern Sie die Einstellungen von Flink log4j-session.properties für Kubernetes-/Yarn-Sitzungen.	Restarts Flink history server.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.	Restarts Flink history server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.	Not available.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Ändert die Werte in der HBase-Umgebung.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.	Not available.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.	This classification should not be reconfigured.
hdfs-env	Ändert die Werte in der HDFS-Umgebung.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei proto-hive-site .xml von HCatalog.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.	Restarts Hive WebHCat server.
hive	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Hive.	Sets configurations to launch Hive LLAP service.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.	Not available.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.	Not available.
hive-env	Ändert die Werte in der Hive-Umgebung.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.	Not available.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log	Not available.
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.	Not available.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.	Not available.



Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hue-ini	Ändert die Werte in der INI-Datei in Hue.	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.	Restarts Hadoop Httpfs service.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.	Not available.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.	Restarts Hadoop-KMS service.
hadoop-kms-java-home	Das KMS-Java-Home von Hadoop ändern	Not available.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.	Not available.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Ändern der Werte in der Hudi-Umgebung.	Not available.
hudi-defaults	Ändern Sie die Werte in der hudi-defaults.conf-Datei in Hudi.	Not available.
iceberg-defaults	Ändern Sie die Werte in der iceberg-defaults.conf-Datei von Iceberg.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
delta-defaults	Ändern Sie die Werte in der delta-defaults.conf-Datei von Delta.	Not available.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.	Not available.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.	Not available.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.	Not available.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.	Not available.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.	Restarts Livy Server.
livy-env	Ändert die Werte in der Livy-Umgebung.	Restarts Livy Server.
livy-log4j2	Ändern Sie die log4j2.properties-Einstellungen für Livy.	Restarts Livy Server.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.	Restarts Hadoop MapReduce-HistoryServer.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
oozie-env	Ändert die Werte in der Oozie-Umgebung.	Restarts Oozie.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.	Restarts Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.	Not available.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.	Not available.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.	Not available.
pig-env	Ändert die Werte in der Pig-Umgebung.	Not available.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.	Restarts Oozie.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.	Not available.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-config	Ändert die Werte in der config.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.	Not available.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.	Not available.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.	Not available.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.	Not available.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.	Not available.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.	Not available.
presto-connector-lakeformation	Ändern Sie die Werte in der lakeformation.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.	Not available.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.	Not available.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.	Not available.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.	Not available.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.	Not available.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.	Not available.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.	Not available.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.	Not available.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.	Not available.
trino-log	Ändern Sie die Werte in der log.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-config	Ändern Sie die Werte in der config.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Ändern Sie die Werte in der Trino-Datei password-authenticator.properties.	Restarts Trino-Server (for Trino)
trino-env	Ändern Sie die Werte in der trino-env.sh-Datei von Trino.	Restarts Trino-Server (for Trino)
trino-node	Ändern Sie die Werte in der node.properties-Datei in Trino.	Not available.
trino-connector-blackhole	Ändern Sie die Werte in der blackhole.properties-Datei in Trino.	Not available.
trino-connector-cassandra	Ändern Sie die Werte in der cassandra.properties-Datei in Trino.	Not available.
trino-connector-delta	Ändern Sie die Werte in der delta.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-hive	Ändern Sie die Werte in der hive.properties-Datei in Trino.	Restarts Trino-Server (for Trino)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-exchange-manager	Ändern Sie die Werte in der <code>exchange-manager.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Ändern Sie die Werte in der <code>iceberg.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-hudi	Ändern Sie die Werte in der <code>hudi.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Ändern Sie die Werte in der <code>jmx.properties</code> -Datei in Trino.	Not available.
trino-connector-kafka	Ändern Sie die Werte in der <code>kafka.properties</code> -Datei in Trino.	Not available.
trino-connector-localfile	Ändern Sie die Werte in der <code>localfile.properties</code> -Datei in Trino.	Not available.
trino-connector-memory	Ändern Sie die Werte in der <code>memory.properties</code> -Datei in Trino.	Not available.
trino-connector-mongodb	Ändern Sie die Werte in der <code>mongodb.properties</code> -Datei in Trino.	Not available.
trino-connector-mysql	Ändern Sie die Werte in der <code>mysql.properties</code> -Datei in Trino.	Not available.
trino-connector-postgresql	Ändern Sie die Werte in der <code>postgresql.properties</code> -Datei in Trino.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-connector-raptor	Ändern Sie die Werte in der raptor.properties-Datei in Trino.	Not available.
trino-connector-redis	Ändern Sie die Werte in der redis.properties-Datei in Trino.	Not available.
trino-connector-redshift	Ändern Sie die Werte in der redshift.properties-Datei in Trino.	Not available.
trino-connector-tpch	Ändern Sie die Werte in der tpch.properties-Datei in Trino.	Not available.
trino-connector-tpcds	Ändern Sie die Werte in der tpcds.properties-Datei in Trino.	Not available.
ranger-kms-dbks-site	Ändert die Werte in der dbks-site.xml-Datei von Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Ändert die Werte in der Ranger KMS-Umgebung.	Restarts Ranger KMS Server.
ranger-kms-logback	Ändern Sie die Werte in der kms-logback.xml-Datei von Ranger KMS.	Not available.
ranger-kms-db-ca	Ändert die Werte für die CA-Datei auf S3 für die MySQL SSL-Verbindung mit Ranger KMS.	Not available.



Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Ändert die Werte in der Spark-Umgebung.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.	Not available.
spark-log4j2	Ändern Sie die Werte in der log4j2.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.	Not available.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.	Not available.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.	Not available.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.	Restart Oozie and HiveServer2.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
yarn-env	Ändert die Werte in der YARN-Umgebung.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.	Restarts Zeppelin.
zeppelin-site	Ändern Sie die Konfigurationseinstellungen in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.	Restarts Zookeeper server.

## 6.15.0 Änderungsprotokoll

### Änderungsprotokoll für Version 6.15.0 und Versionshinweise

Datum	Ereignis	Beschreibung
2023-11-17	Veröffentlichung von Dokumenten	Erste Veröffentlichung der Versionshinweise zu Amazon EMR 6.15.0
17.11.2023-	Die Bereitstellung ist abgeschlossen	<a href="#">Amazon EMR 6.15.0 vollständig in allen unterstützten Regionen bereitgestellt</a>
2023-11-13	Erstversion	Amazon EMR 6.15.0 wurde zuerst in ersten kommerziellen Regionen bereitgestellt

## Amazon-EMR-Version 6.14.0

### 6.14.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Delta](#), [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

## Informationen zur Anwendungsversion

	emr-6.14.0	emr-6.13.0	emr-6.12.0	emr-6.11.1
AWS SDK for Java	1.12.543	1,12,513	1.12,490	1.12,446
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2.12,15	2.12,15	2.12,15	2.12,15
AmazonCloudWatchAgent	-	-	-	-
Delta	2.4.0	2.4.0	2.4.0	2.2.0
Flink	1.17.1	1.17.0	1.17.0	1.16.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4.17	2.4.17	2.4.17	2.4,15
HCatalog	3.1.3	3.1.3	3.1.3	3.1.3
Hadoop	3.3.3	3.3.3	3.3.3	3.3.3
Hive	3.1.3	3.1.3	3.1.3	3.1.3
Hudi	0.13.1-amzn-2	0.13.1-amzn-1	0.13.1-amzn-0	0.13.0-amzn-0
Hue	4.11.0	4.11.0	4.11.0	4.11.0
Iceberg	1.3.1-amzn-0	1.3.0-amzn-1	1.3.0-amzn-0	1.2.0-amzn-0
JupyterEnterpriseGateway	2.6.0	2.6.0	2.6.0	2.6.0
JupyterHub	1.5.0	1.5.0	1.4.1	1.4.1
Livy	0.7.1	0.7.1	0.7.1	0.7.1
MXNet	1.9.1	1.9.1	1.9.1	1.9.1

	emr-6.14.0	emr-6.13.0	emr-6.12.0	emr-6.11.1
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	5.1.3	5.1.3	5.1.3	5.1.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0,281	0,281	0,281	0,279
Spark	3.4.1	3.4.1	3.4.0	3.3.2
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.11.0	2.11.0	2.11.0	2.11.0
Tez	0,102	0.10.2	0.10.2	0.10.2
Trino (PrestoSQL)	422	414	414	410
Zeppelin	0.10.1	0.10.1	0.10.1	0.10.1
ZooKeeper	3.5.10	3.5.10	3.5.10	3.5.10

## 6.14.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon EMR Version 6.14.0. Änderungen beziehen sich auf Version 6.13.0. Informationen zum Zeitplan für die Veröffentlichung finden Sie unter [6.14.0 Änderungsprotokoll](#).

### Neue Features

- Amazon EMR 6.14.0 supports Apache Spark 3.4.1, Apache Spark RAPIDS 23.06.0-amzn-2, Flink 1.17.1, Iceberg 1.3.1, and Trino 422.
- [Amazon EMR Managed Scaling](#) ist jetzt in der Region ap-southeast-3 Asien-Pazifik (Jakarta) für Cluster verfügbar, die Sie mit Amazon EMR 6.14.0 und höher erstellen.

## Änderungen, Verbesserungen und behobene Probleme

- Die Version 6.14.0 optimiert die Protokollverwaltung mit Amazon EMR, das auf Amazon EC2 ausgeführt wird. Infolgedessen können Sie eine leichte Senkung der Speicherkosten für Ihre Cluster-Protokolle feststellen.
- Die Version 6.14.0 verbessert den Skalierungs-Workflow, um verschiedene Core-Instances zu berücksichtigen, deren Größe für ihre Amazon-EBS-Volumes erheblich variiert. Diese Verbesserung gilt nur für Core-Knoten; Herunterskalierungs-Operationen für Aufgabenknoten sind davon nicht betroffen.
- Die Version 6.14.0 verbessert die Interaktion von Amazon EMR mit Open-Source-Anwendungen wie Apache Hadoop YARN ResourceManager and HDFS NameNode. Diese Verbesserung reduziert das Risiko von Betriebsverzögerungen bei der Cluster-Skalierung und verringert Startup-Fehler, die aufgrund von Verbindungsproblemen mit den Open-Source-Anwendungen auftreten.
- Die Version 6.14.0 optimiert die Anwendungsinstallation beim Clusterstart. Dies verbessert die Cluster-Startup-Zeiten für bestimmte Kombinationen von Amazon-EMR-Anwendungen.
- Die Version 6.14.0 behebt ein Problem, bei dem Cluster-Herunterskalierungs-Operationen möglicherweise zum Stillstand kommen, wenn ein Cluster, der in einer VPC mit einer benutzerdefinierten Domain ausgeführt wird, auf einen Core- oder Aufgabenknoten-Neustart stößt.
- Wenn Sie einen Cluster mit der neuesten Patch-Version von Amazon EMR 5.36 oder höher, 6.6 oder höher oder 7.0 oder höher starten, verwendet Amazon EMR die neueste Version von Amazon Linux 2023 oder Amazon Linux 2 für das standardmäßige Amazon EMR-AMI. Weitere Informationen finden Sie unter [Verwenden des standardmäßigen Amazon-Linux-AMI für Amazon EMR](#).

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 223.0	4,14.336	8. März 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			<p>Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost), China (Peking), China (Ningxia), Kanada West (Calgary)</p>

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024131.0	4,14.336	14. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 124.0	4,14.336	7. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsReleaseLabel (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024109.0	4,14.334	24. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 218.0	4,14.330	2. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsReleaseLabel (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202306.0	4,14.330	22. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023116.0	4,14.328	11. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023101.0	4,14.327	17. November 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 4,14.322 906.0		11. September 2023	China (Peking), China (Ningxia)  USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv)

## 6.14.0 Standard-Java-Versionen

Amazon-EMR-Versionen 6.12.0 und höher unterstützen standardmäßig alle Anwendungen mit Amazon Corretto 8, mit Ausnahme von Trino. Für Trino unterstützt Amazon EMR Amazon Corretto 17 ab Amazon-EMR-Version 6.9.0 standardmäßig. Amazon EMR unterstützt auch einige Anwendungen mit Amazon Corretto 11 und 17. Diese Anwendungen werden in der folgenden Tabelle aufgelistet. Wenn Sie die Standard-JVM auf Ihrem Cluster ändern möchten, folgen Sie den Anweisungen unter [Anwendungen für die Verwendung einer bestimmten Java Virtual Machine konfigurieren](#) für jede Anwendung, die auf dem Cluster ausgeführt wird. Sie können nur eine Java-Laufzeit-Version für einen Cluster verwenden. Amazon EMR unterstützt nicht die Ausführung verschiedener Knoten oder Anwendungen auf verschiedenen Laufzeitversionen auf demselben Cluster.

Amazon EMR unterstützt zwar sowohl Amazon Corretto 11 als auch 17 auf Apache Spark, Apache Hadoop und Apache Hive, bei einigen Workloads kann es jedoch zu Leistungseinbußen kommen, wenn Sie diese Versionen von Corretto verwenden. Wir empfehlen Ihnen, Ihre Workloads zu testen, bevor Sie die Standardeinstellungen ändern.

Die folgende Tabelle zeigt die Standard-Java-Versionen für Anwendungen in Amazon EMR 6.14.0:

Anwendung	Java-/Amazon-Corretto-Version (Standard ist fett gedruckt)
Delta	17, 11, 8
Flink	11, 8
Ganglia	8
HBase	11, 8
HCatalog	17, 11, 8
Hadoop	17, 11, 8
Hive	17, 11, 8
Hudi	17, 11, 8
Iceberg	17, 11, 8

Anwendung	Java-/Amazon-Corretto-Version (Standard ist fett gedruckt)
Livy	17, 11, 8
Oozie	17, 11, 8
Phoenix	8
PrestoDB	8
Spark	17, 11, 8
Spark RAPIDS	17, 11, 8
Sqoop	8
Tez	17, 11, 8
Trino	17
Zeppelin	8
Pig	8
Zookeeper	8

## 6.14.0 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die

Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als 2.2-amzn-2 aufgeführt.

Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1.4.2	Amazon SageMaker Spark-SDK
delta	2.4.0	Delta Lake ist ein offenes Tabellenformat für riesige analytische Datensätze
delta-standalone-connectors	0.6.0	Delta-Konnektoren bieten unterschiedliche Laufzeiten, um Delta Lake mit Engines wie Flink, Hive und Presto zu integrieren.
emr-ddb	5.1.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	3.7.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.11.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-notebook-env	1.7.0	Conda Env für EMR-Notebooks, das Jupyter Enterprise Gateway enthält
emr-s3-dist-cp	2.28.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	2.7.0	EMR S3Select-Konnektor



Komponente	Version	Beschreibung
emr-wal-cli	1.1.0	Cli wird für die emrwal list/deletion verwendet.
emrfs	2.59.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.17.1-amzn-0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
flink-jobmanager-config	1.17.1-amzn-0	Verwaltung von Ressourcen auf EMR-Knoten für Apache JobManager Flink.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	3.3.3-amzn-6	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".

Komponente	Version	Beschreibung
hadoop-hdfs-datanode	3.3.3-amzn-6	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	3.3.3-amzn-6	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	3.3.3-amzn-6	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-hdfs-journalnode	3.3.3-amzn-6	HDFS-Service zum Verwalten des Hadoop-Dateisystem journals auf HA-Clustern.
hadoop-httfs-server	3.3.3-amzn-6	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	3.3.3-amzn-6	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	3.3.3-amzn-6	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	3.3.3-amzn-6	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	3.3.3-amzn-6	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.

Komponente	Version	Beschreibung
hadoop-yarn-timeline-server	3.3.3-amzn-6	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	2.4.17-amzn-2	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	2.4.17-amzn-2	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	2.4.17-amzn-2	HBase-Befehlszeilen-Client.
hbase-rest-server	2.4.17-amzn-2	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	2.4.17-amzn-2	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hbase-operator-tools	2.4.17-amzn-2	Reparaturtool für Apache-HBase-Cluster.
hcatalog-client	3.1.3-amzn-7	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	3.1.3-amzn-7	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).

Komponente	Version	Beschreibung
hcatalog-webhcat-server	3.1.3-amzn-7	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	3.1.3-amzn-7	Hive-Befehlszeilen-Client.
hive-hbase	3.1.3-amzn-7	Hive-hbase client.
hive-metastore-server	3.1.3-amzn-7	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	3.1.3-amzn-7	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.13.1-amzn-2	Inkrementelles Verarbeitungs-Framework zur Versorgung der Datenpipeline mit geringer Latenz und hoher Effizienz.
hudi-presto	0.13.1-amzn-2	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hudi-trino	0.13.1-amzn-2	Bündel-Bibliothek zum Ausführen von Trino mit Hudi.
hudi-spark	0.13.1-amzn-2	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.
hue-server	4.11.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.

Komponente	Version	Beschreibung
iceberg	1.3.1-amzn-0	Apache Iceberg ist ein offenes Tabellenformat für sehr große analytische Datensätze
jupyterhub	1.5.0	Multi-User-Server für Jupyter-Notebooks
livy-server	0.7.1-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP- und Reverse-Proxy-Server.
mxnet	1.9.1	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.68+	MariaDB-Datenbankserver.
nvidia-cuda	11,8,0	Nvidia-Treiber und Cuda-Toolkit
oozie-client	5.2.1	Oozie-Befehlszeilen-Client.
oozie-server	5.2.1	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	4.7.0	Open Source Computer Vision Library.
phoenix-library	5.1.3	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-connectors	5.1.3	Apache-Phoenix-Konnektoren für Spark-3

Komponente	Version	Beschreibung
phoenix-query-server	5.1.3	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.281-amzn-2	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.281-amzn-2	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.281-amzn-2	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
trino-coordinator	422-amzn-0	Service zur Annahme von Abfragen und Verwaltung der Abfrageausführung der Trino-Worker.
trino-worker	422-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
trino-client	422-amzn-0	Trino-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Trino-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.

Komponente	Version	Beschreibung
r	4.0.2	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	2.0.0	Apache Ranger Key Management System
spark-client	3.4.1-amzn-1	Spark-Befehlszeilen-Clients.
spark-history-server	3.4.1-amzn-1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	3.4.1-amzn-1	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	3.4.1-amzn-1	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
spark-rapids	23.06.0-amzn-2	Nvidia-Spark-RAPIDS-Plugin, das Apache Spark mit GPUs beschleunigt.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	2.11.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.10.2-amzn-5	Die Tez-YARN-Anwendung und -Bibliotheken.

Komponente	Version	Beschreibung
tez-on-worker	0.10.2-amzn-5	Die Tez-YARN-Anwendung und Bibliotheken für Worker-Knoten.
webserver	2.4.41+	Apache HTTP-Server.
zeppelin-server	0.10.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.5.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.5.10	ZooKeeper Befehlszeilen-Cli-ent.

## 6.14.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Aktionen zur Neukonfiguration treten auf, wenn Sie eine Konfiguration für Instance-Gruppen in einem laufenden Cluster angeben. Amazon EMR initiiert nur Rekonfigurationsaktionen für die Klassifizierungen, die Sie ändern. Weitere Informationen finden Sie unter [Eine Instance-Gruppe in einem laufenden Cluster neu konfigurieren](#).



## emr-6.14.0-Klassifizierungen

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
capacity-scheduler	Ändert die Werte in der capacity-scheduler.xml-Datei in Hadoop.	Restarts the ResourceManager service.
container-executor	Ändern Sie die Werte in der Datei „container-executor.cfg“ Datei von Hadoop YARN.	Not available.
container-log4j	Ändert die Werte in der container-log4j.properties-Datei in Hadoop YARN.	Not available.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Ändern Sie die Docker-bezogenen Einstellungen.	Not available.
emrfs-site	Ändert die EMRFS-Einstellungen.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
		Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegionserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.	Restarts Flink history server.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.	Restarts Flink history server.
flink-log4j-session	Ändern Sie die Einstellungen von Flink log4j-session.properties für Kubernetes-/Yarn-Sitzungen.	Restarts Flink history server.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.	Restarts Flink history server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.	Not available.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Ändert die Werte in der HBase-Umgebung.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.	Not available.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.	This classification should not be reconfigured.
hdfs-env	Ändert die Werte in der HDFS-Umgebung.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei proto-hive-site .xml von HCatalog.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.	Restarts Hive WebHCat server.
hive	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Hive.	Sets configurations to launch Hive LLAP service.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.	Not available.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.	Not available.
hive-env	Ändert die Werte in der Hive-Umgebung.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.	Not available.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log	Not available.
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.	Not available.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hue-ini	Ändert die Werte in der INI-Datei in Hue.	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.	Restarts Hadoop Httpfs service.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.	Not available.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.	Restarts Hadoop-KMS service.
hadoop-kms-java-home	Das KMS-Java-Home von Hadoop ändern	Not available.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.	Not available.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Ändern der Werte in der Hudi-Umgebung.	Not available.
hudi-defaults	Ändern Sie die Werte in der hudi-defaults.conf-Datei in Hudi.	Not available.
iceberg-defaults	Ändern Sie die Werte in der iceberg-defaults.conf-Datei von Iceberg.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
delta-defaults	Ändern Sie die Werte in der delta-defaults.conf-Datei von Delta.	Not available.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.	Not available.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.	Not available.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.	Not available.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.	Not available.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.	Restarts Livy Server.
livy-env	Ändert die Werte in der Livy-Umgebung.	Restarts Livy Server.
livy-log4j2	Ändern Sie die log4j2.properties-Einstellungen für Livy.	Restarts Livy Server.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.	Restarts Hadoop MapReduce-HistoryServer.



Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
oozie-env	Ändert die Werte in der Oozie-Umgebung.	Restarts Oozie.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.	Restarts Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.	Not available.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.	Not available.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.	Not available.
pig-env	Ändert die Werte in der Pig-Umgebung.	Not available.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.	Restarts Oozie.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.	Not available.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-config	Ändert die Werte in der config.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.	Not available.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.	Not available.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.	Not available.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.	Not available.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.	Not available.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.	Not available.
presto-connector-lakeformation	Ändern Sie die Werte in der lakeformation.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.	Not available.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.	Not available.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.	Not available.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.	Not available.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.	Not available.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.	Not available.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.	Not available.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.	Not available.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.	Not available.
trino-log	Ändern Sie die Werte in der log.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-config	Ändern Sie die Werte in der config.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Ändern Sie die Werte in der Trino-Datei password-authenticator.properties.	Restarts Trino-Server (for Trino)
trino-env	Ändern Sie die Werte in der trino-env.sh-Datei von Trino.	Restarts Trino-Server (for Trino)
trino-node	Ändern Sie die Werte in der node.properties-Datei in Trino.	Not available.
trino-connector-blackhole	Ändern Sie die Werte in der blackhole.properties-Datei in Trino.	Not available.
trino-connector-cassandra	Ändern Sie die Werte in der cassandra.properties-Datei in Trino.	Not available.
trino-connector-delta	Ändern Sie die Werte in der delta.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-hive	Ändern Sie die Werte in der hive.properties-Datei in Trino.	Restarts Trino-Server (for Trino)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-exchange-manager	Ändern Sie die Werte in der <code>exchange-manager.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Ändern Sie die Werte in der <code>iceberg.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-hudi	Ändern Sie die Werte in der <code>hudi.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Ändern Sie die Werte in der <code>jmx.properties</code> -Datei in Trino.	Not available.
trino-connector-kafka	Ändern Sie die Werte in der <code>kafka.properties</code> -Datei in Trino.	Not available.
trino-connector-localfile	Ändern Sie die Werte in der <code>localfile.properties</code> -Datei in Trino.	Not available.
trino-connector-memory	Ändern Sie die Werte in der <code>memory.properties</code> -Datei in Trino.	Not available.
trino-connector-mongodb	Ändern Sie die Werte in der <code>mongodb.properties</code> -Datei in Trino.	Not available.
trino-connector-mysql	Ändern Sie die Werte in der <code>mysql.properties</code> -Datei in Trino.	Not available.
trino-connector-postgresql	Ändern Sie die Werte in der <code>postgresql.properties</code> -Datei in Trino.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-connector-raptor	Ändern Sie die Werte in der raptor.properties-Datei in Trino.	Not available.
trino-connector-redis	Ändern Sie die Werte in der redis.properties-Datei in Trino.	Not available.
trino-connector-redshift	Ändern Sie die Werte in der redshift.properties-Datei in Trino.	Not available.
trino-connector-tpch	Ändern Sie die Werte in der tpch.properties-Datei in Trino.	Not available.
trino-connector-tpcds	Ändern Sie die Werte in der tpcds.properties-Datei in Trino.	Not available.
ranger-kms-dbks-site	Ändert die Werte in der dbks-site.xml-Datei von Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Ändert die Werte in der Ranger KMS-Umgebung.	Restarts Ranger KMS Server.
ranger-kms-logback	Ändern Sie die Werte in der kms-logback.xml-Datei von Ranger KMS.	Not available.
ranger-kms-db-ca	Ändert die Werte für die CA-Datei auf S3 für die MySQL SSL-Verbindung mit Ranger KMS.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Ändert die Werte in der Spark-Umgebung.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.	Not available.
spark-log4j2	Ändern Sie die Werte in der log4j2.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.	Not available.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.	Not available.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.	Not available.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.	Restart Oozie and HiveServer2.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
yarn-env	Ändert die Werte in der YARN-Umgebung.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.	Restarts Zeppelin.
zeppelin-site	Ändern Sie die Konfigurationseinstellungen in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.	Restarts Zookeeper server.



## 6.14.0 Änderungsprotokoll

Änderungsprotokoll für Version 6.14.0 und Versionshinweise

Datum	Ereignis	Beschreibung
2023-11-02	Die Bereitstellung ist abgeschlossen	<a href="#">Amazon EMR 6.14.0 vollständig in allen unterstützten Regionen bereitgestellt</a>
2023-10-10	Veröffentlichung von Dokumenten	Erste Veröffentlichung der Versionshinweise zu Amazon EMR 6.14.0
2023-10-04	Erstversion	Amazon EMR 6.14.0 wurde zuerst in ersten kommerziellen Regionen bereitgestellt

## Amazon-EMR-Version 6.13.0

### 6.13.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Delta](#), [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

## Informationen zur Anwendungsversion

	emr-6.13.0	emr-6.12.0	emr-6.11.1	emr-6.11.0
AWS SDK for Java	1.12.513	1.12,490	1.12,446	1.12,446
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2.12,15	2.12,15	2.12,15	2.12,15
AmazonCloudWatchAgent	-	-	-	-
Delta	2.4.0	2.4.0	2.2.0	2.2.0
Flink	1.17.0	1.17.0	1.16.0	1.16.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4.17	2.4.17	2.4,15	2.4,15
HCatalog	3.1.3	3.1.3	3.1.3	3.1.3
Hadoop	3.3.3	3.3.3	3.3.3	3.3.3
Hive	3.1.3	3.1.3	3.1.3	3.1.3
Hudi	0.13.1-amzn-1	0.13.1-amzn-0	0.13.0-amzn-0	0.13.0-amzn-0
Hue	4.11.0	4.11.0	4.11.0	4.11.0
Iceberg	1.3.0-amzn-1	1.3.0-amzn-0	1.2.0-amzn-0	1.2.0-amzn-0
JupyterEnterpriseGateway	2.6.0	2.6.0	2.6.0	2.6.0
JupyterHub	1.5.0	1.4.1	1.4.1	1.4.1
Livy	0.7.1	0.7.1	0.7.1	0.7.1
MXNet	1.9.1	1.9.1	1.9.1	1.9.1

	emr-6.13.0	emr-6.12.0	emr-6.11.1	emr-6.11.0
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	5.1.3	5.1.3	5.1.2	5.1.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0,281	0,281	0,279	0,279
Spark	3.4.1	3.4.0	3.3.2	3.3.2
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.11.0	2.11.0	2.11.0	2.11.0
Tez	0,102	0.10.2	0.10.2	0.10.2
Trino (PrestoSQL)	414	414	410	410
Zeppelin	0.10.1	0.10.1	0.10.1	0.10.1
ZooKeeper	3.5.10	3.5.10	3.5.10	3.5.10

## 6.13.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 6.13.0. Änderungen beziehen sich auf Version 6.12.0. Informationen zum Zeitplan für die Veröffentlichung finden Sie unter [6.13.0 Änderungsprotokoll](#).

### Neue Features

- Amazon EMR 6.13.0 supports Apache Spark 3.4.1, Apache Spark RAPIDS 23.06.0-amzn-1, CUDA Toolkit 11.8.0, and JupyterHub 1.5.0.

## Änderungen, Verbesserungen und behobene Probleme

- Die Version 6.13.0 verbessert den Amazon-EMR-Protokollverwaltungs-Daemon, um sicherzustellen, dass alle Protokolle in regelmäßigen Abständen auf Amazon S3 hochgeladen werden, wenn ein Befehl zur Clusterbeendigung ausgegeben wird. Dies ermöglicht schnellere Clusterbeendigungen.
- Die Version 6.13.0 verbessert die Protokollverwaltungsfunktionen von Amazon EMR, um einen konsistenten und zeitnahen Upload aller Protokolldateien auf Amazon S3 sicherzustellen. Davon profitieren vor allem langlaufende EMR-Cluster.
- Wenn Sie einen Cluster mit der neuesten Patch-Version von Amazon EMR 5.36 oder höher, 6.6 oder höher oder 7.0 oder höher starten, verwendet Amazon EMR die neueste Version von Amazon Linux 2023 oder Amazon Linux 2 für das standardmäßige Amazon EMR-AMI. Weitere Informationen finden Sie unter [Verwenden des standardmäßigen Amazon-Linux-AMI für Amazon EMR](#).

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 223.0	4,14.336	8. März 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost), China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024131.0	4,14.336	14. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 124.0	4,14.336	7. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024109.0	4,14.334	24. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 218.0	4,14.330	2. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202306.0	4,14.330	22. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 116.0	4,14.328	11. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023101.0	4,14.327	16. November 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023020.1	4,14.326	7. November 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023012.1	4,14.326	26. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 926.0	4,14.322	19. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)
2.0.2023 906.0	4,14.322	04. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 808,0	4,14.320	24. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv)

## 6.13.0 Standard-Java-Versionen

Amazon-EMR-Versionen 6.12.0 und höher unterstützen standardmäßig alle Anwendungen mit Amazon Corretto 8, mit Ausnahme von Trino. Für Trino unterstützt Amazon EMR Amazon Corretto 17 ab Amazon-EMR-Version 6.9.0 standardmäßig. Amazon EMR unterstützt auch einige Anwendungen mit Amazon Corretto 11 und 17. Diese Anwendungen werden in der folgenden Tabelle aufgelistet. Wenn Sie die Standard-JVM auf Ihrem Cluster ändern möchten, folgen Sie den Anweisungen unter [Anwendungen für die Verwendung einer bestimmten Java Virtual Machine konfigurieren](#) für jede Anwendung, die auf dem Cluster ausgeführt wird. Sie können nur eine Java-Laufzeit-Version für einen Cluster verwenden. Amazon EMR unterstützt nicht die Ausführung verschiedener Knoten oder Anwendungen auf verschiedenen Laufzeitversionen auf demselben Cluster.

Amazon EMR unterstützt zwar sowohl Amazon Corretto 11 als auch 17 auf Apache Spark, Apache Hadoop und Apache Hive, bei einigen Workloads kann es jedoch zu Leistungseinbußen kommen, wenn Sie diese Versionen von Corretto verwenden. Wir empfehlen Ihnen, Ihre Workloads zu testen, bevor Sie die Standardeinstellungen ändern.

Die folgende Tabelle zeigt die Standard-Java-Versionen für Anwendungen in Amazon EMR 6.13.0:

Anwendung	Java-/Amazon-Corretto-Version (Standard ist fett gedruckt)
Delta	17, 11, 8
Flink	11, 8
Ganglia	8
HBase	11, 8
HCatalog	17, 11, 8
Hadoop	17, 11, 8
Hive	17, 11, 8
Hudi	17, 11, 8
Iceberg	17, 11, 8

Anwendung	Java-/Amazon-Corretto-Version (Standard ist fett gedruckt)
Livy	17, 11, 8
Oozie	17, 11, 8
Phoenix	8
PrestoDB	8
Spark	17, 11, 8
Spark RAPIDS	17, 11, 8
Sqoop	8
Tez	17, 11, 8
Trino	17
Zeppelin	8
Pig	8
Zookeeper	8

### 6.13.0 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die

Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als 2.2-amzn-2 aufgeführt.

Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1.4.2	Amazon SageMaker Spark-SDK
delta	2.4.0	Delta Lake ist ein offenes Tabellenformat für riesige analytische Datensätze
delta-standalone-connectors	0.6.0	Delta-Konnektoren bieten unterschiedliche Laufzeiten, um Delta Lake mit Engines wie Flink, Hive und Presto zu integrieren.
emr-ddb	5.1.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	3.6.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.10.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-notebook-env	1.7.0	Conda Env für EMR-Notebooks, das Jupyter Enterprise Gateway enthält
emr-s3-dist-cp	2.27.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	2.6.0	EMR S3Select-Konnektor

Komponente	Version	Beschreibung
emr-wal-cli	1.1.0	Cli wird für die emrwal list/deletion verwendet.
emrfs	2,58,0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.17.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
flink-jobmanager-config	1.17.0	Verwaltung von Ressourcen auf EMR-Knoten für Apache JobManager Flink.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	3.3.3-amzn-5	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".

Komponente	Version	Beschreibung
hadoop-hdfs-datanode	3.3.3-amzn-5	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	3.3.3-amzn-5	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	3.3.3-amzn-5	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-hdfs-journalnode	3.3.3-amzn-5	HDFS-Service zum Verwalten des Hadoop-Dateisystem journals auf HA-Clustern.
hadoop-httfs-server	3.3.3-amzn-5	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	3.3.3-amzn-5	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	3.3.3-amzn-5	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	3.3.3-amzn-5	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	3.3.3-amzn-5	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.

Komponente	Version	Beschreibung
hadoop-yarn-timeline-server	3.3.3-amzn-5	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	2.4.17-amzn-1	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	2.4.17-amzn-1	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	2.4.17-amzn-1	HBase-Befehlszeilen-Client.
hbase-rest-server	2.4.17-amzn-1	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	2.4.17-amzn-1	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hbase-operator-tools	2.4.17-amzn-1	Reparaturtool für Apache-HBase-Cluster.
hcatalog-client	3.1.3-amzn-6	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	3.1.3-amzn-6	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).



Komponente	Version	Beschreibung
hcatalog-webhcat-server	3.1.3-amzn-6	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	3.1.3-amzn-6	Hive-Befehlszeilen-Client.
hive-hbase	3.1.3-amzn-6	Hive-hbase client.
hive-metastore-server	3.1.3-amzn-6	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	3.1.3-amzn-6	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.13.1-amzn-1	Inkrementelles Verarbeitungs-Framework zur Versorgung der Datenpipeline mit geringer Latenz und hoher Effizienz.
hudi-presto	0.13.1-amzn-1	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hudi-trino	0.13.1-amzn-1	Bündel-Bibliothek zum Ausführen von Trino mit Hudi.
hudi-spark	0.13.1-amzn-1	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.
hue-server	4.11.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.

Komponente	Version	Beschreibung
iceberg	1.3.0-amzn-1	Apache Iceberg ist ein offenes Tabellenformat für sehr große analytische Datensätze
jupyterhub	1.5.0	Multi-User-Server für Jupyter-Notebooks
livy-server	0.7.1-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP- und Reverse-Proxy-Server.
mxnet	1.9.1	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.68+	MariaDB-Datenbankserver.
nvidia-cuda	11,8,0	Nvidia-Treiber und Cuda-Toolkit
oozie-client	5.2.1	Oozie-Befehlszeilen-Client.
oozie-server	5.2.1	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	4.7.0	Open Source Computer Vision Library.
phoenix-library	5.1.3	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-connectors	5.1.3	Apache-Phoenix-Konnektoren für Spark-3

Komponente	Version	Beschreibung
phoenix-query-server	5.1.3	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.281-amzn-1	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.281-amzn-1	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.281-amzn-1	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
trino-coordinator	414-amzn-1	Service zur Annahme von Abfragen und Verwaltung der Abfrageausführung der Trino-Worker.
trino-worker	414-amzn-1	Service für das Ausführen von Teilen einer Abfrage.
trino-client	414-amzn-1	Trino-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Trino-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.

Komponente	Version	Beschreibung
r	4.0.2	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	2.0.0	Apache Ranger Key Management System
spark-client	3.4.1-amzn-0	Spark-Befehlszeilen-Clients.
spark-history-server	3.4.1-amzn-0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	3.4.1-amzn-0	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	3.4.1-amzn-0	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
spark-rapids	23.06.0-amzn-1	Nvidia-Spark-RAPIDS-Plugin, das Apache Spark mit GPUs beschleunigt.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	2.11.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.10.2-amzn-4	Die Tez-YARN-Anwendung und -Bibliotheken.

Komponente	Version	Beschreibung
tez-on-worker	0.10.2-amzn-4	Die Tez-YARN-Anwendung und Bibliotheken für Worker-Knoten.
webserver	2.4.41+	Apache HTTP-Server.
zeppelin-server	0.10.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.5.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.5.10	ZooKeeper Befehlszeilen-Cli-ent.

### 6.13.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Aktionen zur Neukonfiguration treten auf, wenn Sie eine Konfiguration für Instance-Gruppen in einem laufenden Cluster angeben. Amazon EMR initiiert nur Rekonfigurationsaktionen für die Klassifizierungen, die Sie ändern. Weitere Informationen finden Sie unter [Eine Instance-Gruppe in einem laufenden Cluster neu konfigurieren](#).

## emr-6.13.0-Klassifizierungen

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
capacity-scheduler	Ändert die Werte in der capacity-scheduler.xml-Datei in Hadoop.	Restarts the ResourceManager service.
container-executor	Ändern Sie die Werte in der Datei „container-executor.cfg“ Datei von Hadoop YARN.	Not available.
container-log4j	Ändert die Werte in der container-log4j.properties-Datei in Hadoop YARN.	Not available.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Ändern Sie die Docker-bezogenen Einstellungen.	Not available.
emrfs-site	Ändert die EMRFS-Einstellungen.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
		Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegionserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.	Restarts Flink history server.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.	Restarts Flink history server.
flink-log4j-session	Ändern Sie die Einstellungen von Flink log4j-session.properties für Kubernetes-/Yarn-Sitzungen.	Restarts Flink history server.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.	Restarts Flink history server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.	Not available.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.	Not available.



Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Ändert die Werte in der HBase-Umgebung.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.	Not available.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.	This classification should not be reconfigured.
hdfs-env	Ändert die Werte in der HDFS-Umgebung.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei proto-hive-site .xml von HCatalog.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.	Restarts Hive WebHCat server.
hive	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Hive.	Sets configurations to launch Hive LLAP service.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.	Not available.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.	Not available.
hive-env	Ändert die Werte in der Hive-Umgebung.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.	Not available.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log	Not available.
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.	Not available.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hue-ini	Ändert die Werte in der INI-Datei in Hue.	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.	Restarts Hadoop Httpfs service.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.	Not available.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.	Restarts Hadoop-KMS service.
hadoop-kms-java-home	Das KMS-Java-Home von Hadoop ändern	Not available.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.	Not available.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Ändern der Werte in der Hudi-Umgebung.	Not available.
hudi-defaults	Ändern Sie die Werte in der hudi-defaults.conf-Datei in Hudi.	Not available.
iceberg-defaults	Ändern Sie die Werte in der iceberg-defaults.conf-Datei von Iceberg.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
delta-defaults	Ändern Sie die Werte in der delta-defaults.conf-Datei von Delta.	Not available.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.	Not available.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.	Not available.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.	Not available.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.	Not available.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.	Restarts Livy Server.
livy-env	Ändert die Werte in der Livy-Umgebung.	Restarts Livy Server.
livy-log4j2	Ändern Sie die log4j2.properties-Einstellungen für Livy.	Restarts Livy Server.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.	Restarts Hadoop MapReduce-HistoryServer.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
oozie-env	Ändert die Werte in der Oozie-Umgebung.	Restarts Oozie.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.	Restarts Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.	Not available.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.	Not available.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.	Not available.
pig-env	Ändert die Werte in der Pig-Umgebung.	Not available.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.	Restarts Oozie.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.	Not available.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-config	Ändert die Werte in der config.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.	Not available.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.	Not available.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.	Not available.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.	Not available.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.	Not available.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.	Not available.
presto-connector-lakeformation	Ändern Sie die Werte in der lakeformation.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.	Not available.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.	Not available.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.	Not available.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.	Not available.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.	Not available.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.	Not available.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.	Not available.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.	Not available.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.	Not available.



Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.	Not available.
trino-log	Ändern Sie die Werte in der log.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-config	Ändern Sie die Werte in der config.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Ändern Sie die Werte in der Trino-Datei password-authenticator.properties.	Restarts Trino-Server (for Trino)
trino-env	Ändern Sie die Werte in der trino-env.sh-Datei von Trino.	Restarts Trino-Server (for Trino)
trino-node	Ändern Sie die Werte in der node.properties-Datei in Trino.	Not available.
trino-connector-blackhole	Ändern Sie die Werte in der blackhole.properties-Datei in Trino.	Not available.
trino-connector-cassandra	Ändern Sie die Werte in der cassandra.properties-Datei in Trino.	Not available.
trino-connector-delta	Ändern Sie die Werte in der delta.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-hive	Ändern Sie die Werte in der hive.properties-Datei in Trino.	Restarts Trino-Server (for Trino)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-exchange-manager	Ändern Sie die Werte in der <code>exchange-manager.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Ändern Sie die Werte in der <code>iceberg.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-hudi	Ändern Sie die Werte in der <code>hudi.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Ändern Sie die Werte in der <code>jmx.properties</code> -Datei in Trino.	Not available.
trino-connector-kafka	Ändern Sie die Werte in der <code>kafka.properties</code> -Datei in Trino.	Not available.
trino-connector-localfile	Ändern Sie die Werte in der <code>localfile.properties</code> -Datei in Trino.	Not available.
trino-connector-memory	Ändern Sie die Werte in der <code>memory.properties</code> -Datei in Trino.	Not available.
trino-connector-mongodb	Ändern Sie die Werte in der <code>mongodb.properties</code> -Datei in Trino.	Not available.
trino-connector-mysql	Ändern Sie die Werte in der <code>mysql.properties</code> -Datei in Trino.	Not available.
trino-connector-postgresql	Ändern Sie die Werte in der <code>postgresql.properties</code> -Datei in Trino.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-connector-raptor	Ändern Sie die Werte in der raptor.properties-Datei in Trino.	Not available.
trino-connector-redis	Ändern Sie die Werte in der redis.properties-Datei in Trino.	Not available.
trino-connector-redshift	Ändern Sie die Werte in der redshift.properties-Datei in Trino.	Not available.
trino-connector-tpch	Ändern Sie die Werte in der tpch.properties-Datei in Trino.	Not available.
trino-connector-tpcds	Ändern Sie die Werte in der tpcds.properties-Datei in Trino.	Not available.
ranger-kms-dbks-site	Ändert die Werte in der dbks-site.xml-Datei von Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Ändert die Werte in der Ranger KMS-Umgebung.	Restarts Ranger KMS Server.
ranger-kms-logback	Ändern Sie die Werte in der kms-logback.xml-Datei von Ranger KMS.	Not available.
ranger-kms-db-ca	Ändert die Werte für die CA-Datei auf S3 für die MySQL SSL-Verbindung mit Ranger KMS.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Ändert die Werte in der Spark-Umgebung.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.	Not available.
spark-log4j2	Ändern Sie die Werte in der log4j2.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.	Not available.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.	Not available.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.	Not available.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.	Restart Oozie and HiveServer2.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
yarn-env	Ändert die Werte in der YARN-Umgebung.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.	Restarts Zeppelin.
zeppelin-site	Ändern Sie die Konfigurationseinstellungen in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.	Restarts Zookeeper server.

## 6.13.0 Änderungsprotokoll

Änderungsprotokoll für Version 6.13.0 und Versionshinweise

Datum	Ereignis	Beschreibung
2023-09-23	Die Bereitstellung ist abgeschlossen	<a href="#">Amazon EMR 6.13.0 vollständig in allen unterstützten Regionen bereitgestellt</a>
2023-09-12	Veröffentlichung von Dokumenten	Erste Veröffentlichung der Versionshinweise zu Amazon EMR 6.13.0
2023-09-01	Erstversion	Amazon EMR 6.13.0 wurde zuerst in ersten kommerziellen Regionen bereitgestellt

## Amazon-EMR-Version 6.12.0

### 6.12.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Delta](#), [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

## Informationen zur Anwendungsversion

	emr-6.12.0	emr-6.11.1	emr-6.11.0	emr-6.10.1
AWS SDK for Java	1.12.490	1.12,446	1.12,446	1.12,397
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2.12,15	2.12,15	2.12,15	2.12,15
AmazonCloudWatchAgent	-	-	-	-
Delta	2.4.0	2.2.0	2.2.0	2.2.0
Flink	1.17.0	1.16.0	1.16.0	1.16.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4.17	2.4,15	2.4,15	2.4,15
HCatalog	3.1.3	3.1.3	3.1.3	3.1.3
Hadoop	3.3.3	3.3.3	3.3.3	3.3.3
Hive	3.1.3	3.1.3	3.1.3	3.1.3
Hudi	0.13.1-amzn-0	0.13.0-amzn-0	0.13.0-amzn-0	0.12.2-amzn-0
Hue	4.11.0	4.11.0	4.11.0	4.10.0
Iceberg	1.3.0-amzn-0	1.2.0-amzn-0	1.2.0-amzn-0	1.1.0-amzn-0
JupyterEnterpriseGateway	2.6.0	2.6.0	2.6.0	2.6.0
JupyterHub	1.4.1	1.4.1	1.4.1	1.5.0
Livy	0.7.1	0.7.1	0.7.1	0.7.1
MXNet	1.9.1	1.9.1	1.9.1	1.9.1

	emr-6.12.0	emr-6.11.1	emr-6.11.0	emr-6.10.1
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	5.1.3	5.1.2	5.1.2	5.1.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0,281	0,279	0,279	0,278
Spark	3.4.0	3.3.2	3.3.2	3.3.1
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.11.0	2.11.0	2.11.0	2.11.0
Tez	0,102	0.10.2	0.10.2	0.10.2
Trino (PrestoSQL)	414	410	410	403
Zeppelin	0.10.1	0.10.1	0.10.1	0.10.1
ZooKeeper	3.5.10	3.5.10	3.5.10	3.5.10

## 6.12.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 6.12.0. Änderungen beziehen sich auf Version 6.11.0. Informationen zum Zeitplan für die Veröffentlichung finden Sie unter [6.12.0 Änderungsprotokoll](#).

### Neue Features

- Amazon EMR 6.12.0 supports Apache Spark 3.4.0, Apache Spark RAPIDS 23.06.0-amzn-0, CUDA 11.8.0, Apache Hudi 0.13.1-amzn-0, Apache Iceberg 1.3.0-amzn-0, Trino 414, and PrestoDB 0.281.
- Amazon EMR-Versionen 6.12.0 und höher unterstützen die LDAP-Integration mit Apache Livy, Apache Hive bis HiveServer 2 (HS2), Trino, Presto und Hue. Sie können Apache Spark und



Apache Hadoop auch auf einem EMR-Cluster installieren, der 6.12.0 oder höher verwendet, und sie für die Verwendung von LDAP konfigurieren. Weitere Informationen finden Sie unter [Verwenden von Active-Directory- oder LDAP-Servern für die Authentifizierung mit Amazon EMR](#).

## Änderungen, Verbesserungen und behobene Probleme

- Amazon-EMR-Versionen 6.12.0 und höher bieten Java-11-Laufzeitunterstützung für Flink. Weitere Informationen finden Sie unter [Flink für die Ausführung mit Java 11 konfigurieren](#).
- Die Version 6.12.0 fügt dem Cluster-Skalierungs-Workflow für EMR-Cluster, auf denen Presto oder Trino ausgeführt werden, einen neuen Wiederholungsmechanismus hinzu. Diese Verbesserung verringert das Risiko, dass die Clustergrößenänderung aufgrund eines einzigen fehlgeschlagenen Größenänderungsvorgangs auf unbestimmte Zeit zum Stillstand kommt. Sie verbessert auch die Clusterauslastung, da Ihr Cluster schneller hoch- und herunterskaliert wird.
- Die Version 6.12.0 behebt ein Problem, bei dem Cluster-Herunterskalierungs-Operationen zum Stillstand kommen können, wenn ein Core-Knoten, der einer ordnungsgemäßen Außerbetriebnahme unterzogen wird, aus irgendeinem Grund nicht mehr funktionsfähig ist, bevor er vollständig außer Betrieb genommen wird.
- Die Version 6.12.0 verbessert die Cluster-Herunterskalierungs-Logik, sodass Ihr Cluster nicht versucht, die Core-Knoten unter die für den Cluster eingestellte HDFS-Replikationsfaktor-Einstellung herunterzuskalieren. Dies entspricht Ihren Anforderungen an die Datenredundanz und verringert die Wahrscheinlichkeit, dass ein Skalierungsvorgang zum Stillstand kommt.
- Die Version 6.12.0 verbessert die Leistung und Effizienz des Health Monitoring Service für Amazon EMR, indem die Geschwindigkeit erhöht wird, mit der Statusänderungen für Instances protokolliert werden. Diese Verbesserung verringert das Risiko einer Leistungseinbuße bei Cluster-Knoten, auf denen mehrere benutzerdefinierte Client-Tools oder Drittanbieteranwendungen ausgeführt werden.
- Die Version 6.12.0 verbessert die Leistung des On-Cluster-Protokollverwaltungs-Daemons für Amazon EMR. Daher ist die Wahrscheinlichkeit einer Leistungsverschlechterung bei EMR-Clustern, die Schritte mit hoher Parallelität ausführen, geringer.
- Mit der Amazon-EMR-Version 6.12.0 wurde der Protokoll-Management-Daemon aktualisiert, um alle Protokolle, die aktiv verwendet werden, mit offenen Datei-Handles auf dem lokalen Instance-Speicher und die zugehörigen Prozesse zu identifizieren. Dieses Upgrade stellt sicher, dass Amazon EMR die Dateien ordnungsgemäß löscht und Speicherplatz zurückgewinnt, nachdem die Protokolle in Amazon S3 archiviert wurden.
- Die Version 6.12.0 beinhaltet eine Erweiterung des Protokollverwaltungs-Daemons, die leere, unbenutzte Schrittverzeichnisse im lokalen Cluster-Dateisystem löscht. Eine zu große Anzahl

leerer Verzeichnisse kann die Leistung der Amazon-EMR-Daemons beeinträchtigen und zu einer Überauslastung der Festplatte führen.

- Die Version 6.12.0 ermöglicht die Protokollrotation für YARN Timeline Server-Protokolle. Dadurch werden Szenarien mit übermäßiger Festplattenauslastung minimiert, insbesondere bei Clustern mit langer Laufzeit.
- Die Standardgröße des Root-Volumes wurde in Amazon EMR 6.10.0 und höher auf 15 GB erhöht. Frühere Versionen haben eine Standardgröße für das Root-Volume von 10 GB.
- Wenn Sie einen Cluster mit der neuesten Patch-Version von Amazon EMR 5.36 oder höher, 6.6 oder höher oder 7.0 oder höher starten, verwendet Amazon EMR die neueste Version von Amazon Linux 2023 oder Amazon Linux 2 für das standardmäßige Amazon EMR-AMI. Weitere Informationen finden Sie unter [Verwenden des standardmäßigen Amazon-Linux-AMI für Amazon EMR](#).

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 223.0	4,14.336	8. März 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			<p>Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost), China (Peking), China (Ningxia), Kanada West (Calgary)</p>

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024131.0	4,14.336	14. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 124.0	4,14.336	7. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsReleaseLabel (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024109.0	4,14.334	24. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 218.0	4,14.330	2. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202306.0	4,14.330	22. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 116.0	4,14.328	11. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023101.0	4,14.327	16. November 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023020.1	4,14.326	7. November 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023012.1	4,14.326	26. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 926.0	4,14.322	19. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)
2.0.2023 906.0	4,14.322	04. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 822.0	4,14.322	30. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv)



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.20230808,0	4,14.320	24. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 727.0	4,14.320	14. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 719,0	4,14.320	02. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 628.0	4,14.318	12. Juli 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral)

## 6.12.0 Standard-Java-Versionen

Amazon-EMR-Versionen 6.12.0 und höher unterstützen standardmäßig alle Anwendungen mit Amazon Corretto 8, mit Ausnahme von Trino. Für Trino unterstützt Amazon EMR Amazon Corretto 17

ab Amazon-EMR-Version 6.9.0 standardmäßig. Amazon EMR unterstützt auch einige Anwendungen mit Amazon Corretto 11 und 17. Diese Anwendungen werden in der folgenden Tabelle aufgelistet. Wenn Sie die Standard-JVM auf Ihrem Cluster ändern möchten, folgen Sie den Anweisungen unter [Anwendungen für die Verwendung einer bestimmten Java Virtual Machine konfigurieren](#) für jede Anwendung, die auf dem Cluster ausgeführt wird. Sie können nur eine Java-Laufzeit-Version für einen Cluster verwenden. Amazon EMR unterstützt nicht die Ausführung verschiedener Knoten oder Anwendungen auf verschiedenen Laufzeitversionen auf demselben Cluster.

Amazon EMR unterstützt zwar sowohl Amazon Corretto 11 als auch 17 auf Apache Spark, Apache Hadoop und Apache Hive, bei einigen Workloads kann es jedoch zu Leistungseinbußen kommen, wenn Sie diese Versionen von Corretto verwenden. Wir empfehlen Ihnen, Ihre Workloads zu testen, bevor Sie die Standardeinstellungen ändern.

Die folgende Tabelle zeigt die Standard-Java-Versionen für Anwendungen in Amazon EMR 6.12.0:

Anwendung	Java-/Amazon-Corretto-Version (Standard ist fett gedruckt)
Delta	17, 11, 8
Flink	11, 8
Ganglia	8
HBase	11, 8
HCatalog	17, 11, 8
Hadoop	17, 11, 8
Hive	17, 11, 8
Hudi	17, 11, 8
Iceberg	17, 11, 8
Livy	17, 11, 8
Oozie	17, 11, 8
Phoenix	8

Anwendung	Java-/Amazon-Corretto-Version (Standard ist fett gedruckt)
PrestoDB	8
Spark	17, 11, 8
Spark RAPIDS	17, 11, 8
Sqoop	8
Tez	17, 11, 8
Trino	17
Zeppelin	8
Pig	8
Zookeeper	8

## 6.12.0 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1.4.2	Amazon SageMaker Spark-SDK
delta	2.4.0	Delta Lake ist ein offenes Tabellenformat für riesige analytische Datensätze
delta-standalone-connectors	0.6.0	Delta-Konnektoren bieten unterschiedliche Laufzeiten, um Delta Lake mit Engines wie Flink, Hive und Presto zu integrieren.
emr-ddb	5.1.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	3.5.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.9.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-notebook-env	1.7.0	Conda Env für EMR-Notebooks, das Jupyter Enterprise Gateway enthält
emr-s3-dist-cp	2.26.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	2.5.0	EMR S3Select-Konnektor
emr-wal-cli	1.1.0	Cli wird für die emrwal list/deletion verwendet.

Komponente	Version	Beschreibung
emrfs	2,57,0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.17.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
flink-jobmanager-config	1.17.0	Verwaltung von Ressourcen auf EMR-Knoten für Apache JobManager Flink.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	3.3.3-amzn-4	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	3.3.3-amzn-4	HDFS-Service auf Knotenebene zum Speichern von Blöcken.



Komponente	Version	Beschreibung
hadoop-hdfs-library	3.3.3-amzn-4	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	3.3.3-amzn-4	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-hdfs-journalnode	3.3.3-amzn-4	HDFS-Service zum Verwalten des Hadoop-Dateisystem journals auf HA-Clustern.
hadoop-httfs-server	3.3.3-amzn-4	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	3.3.3-amzn-4	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	3.3.3-amzn-4	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	3.3.3-amzn-4	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	3.3.3-amzn-4	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.

Komponente	Version	Beschreibung
hadoop-yarn-timeline-server	3.3.3-amzn-4	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	2.4.17-amzn-0	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	2.4.17-amzn-0	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	2.4.17-amzn-0	HBase-Befehlszeilen-Client.
hbase-rest-server	2.4.17-amzn-0	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	2.4.17-amzn-0	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hbase-operator-tools	2.4.17-amzn-0	Reparaturtool für Apache-HBase-Cluster.
hcatalog-client	3.1.3-amzn-5	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	3.1.3-amzn-5	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).

Komponente	Version	Beschreibung
hcatalog-webhcat-server	3.1.3-amzn-5	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	3.1.3-amzn-5	Hive-Befehlszeilen-Client.
hive-hbase	3.1.3-amzn-5	Hive-hbase client.
hive-metastore-server	3.1.3-amzn-5	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	3.1.3-amzn-5	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.13.1-amzn-0	Inkrementelles Verarbeitungs-Framework zur Versorgung der Datenpipeline mit geringer Latenz und hoher Effizienz.
hudi-presto	0.13.1-amzn-0	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hudi-trino	0.13.1-amzn-0	Bündel-Bibliothek zum Ausführen von Trino mit Hudi.
hudi-spark	0.13.1-amzn-0	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.
hue-server	4.11.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.

Komponente	Version	Beschreibung
iceberg	1.3.0-amzn-0	Apache Iceberg ist ein offenes Tabellenformat für sehr große analytische Datensätze
jupyterhub	1.4.1	Multi-User-Server für Jupyter-Notebooks
livy-server	0.7.1-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP- und Reverse-Proxy-Server.
mxnet	1.9.1	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.68+	MariaDB-Datenbankserver.
nvidia-cuda	11,8,0	Nvidia-Treiber und Cuda-Toolkit
oozie-client	5.2.1	Oozie-Befehlszeilen-Client.
oozie-server	5.2.1	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	4.7.0	Open Source Computer Vision Library.
phoenix-library	5.1.3	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-connectors	5.1.3	Apache-Phoenix-Konnektoren für Spark-3

Komponente	Version	Beschreibung
phoenix-query-server	5.1.3	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.281-amzn-0	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.281-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.281-amzn-0	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
trino-coordinator	414-amzn-0	Service zur Annahme von Abfragen und Verwaltung der Abfrageausführung der Trino-Worker.
trino-worker	414-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
trino-client	414-amzn-0	Trino-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Trino-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.

Komponente	Version	Beschreibung
r	4.0.2	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	2.0.0	Apache Ranger Key Management System
spark-client	3.4.0-amzn-0	Spark-Befehlszeilen-Clients.
spark-history-server	3.4.0-amzn-0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	3.4.0-amzn-0	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	3.4.0-amzn-0	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
spark-rapids	23.06.0-amzn-0	Nvidia-Spark-RAPIDS-Plugin, das Apache Spark mit GPUs beschleunigt.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	2.11.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.10.2-amzn-3	Die Tez-YARN-Anwendung und -Bibliotheken.

Komponente	Version	Beschreibung
tez-on-worker	0.10.2-amzn-3	Die Tez-YARN-Anwendung und Bibliotheken für Worker-Knoten.
webserver	2.4.41+	Apache HTTP-Server.
zeppelin-server	0.10.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.5.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.5.10	ZooKeeper Befehlszeilen-Cli-ent.

## 6.12.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Aktionen zur Neukonfiguration treten auf, wenn Sie eine Konfiguration für Instance-Gruppen in einem laufenden Cluster angeben. Amazon EMR initiiert nur Rekonfigurationsaktionen für die Klassifizierungen, die Sie ändern. Weitere Informationen finden Sie unter [Eine Instance-Gruppe in einem laufenden Cluster neu konfigurieren](#).

## emr-6.12.0-Klassifizierungen

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
capacity-scheduler	Ändert die Werte in der capacity-scheduler.xml-Datei in Hadoop.	Restarts the ResourceManager service.
container-executor	Ändern Sie die Werte in der Datei „container-executor.cfg“ Datei von Hadoop YARN.	Not available.
container-log4j	Ändert die Werte in der container-log4j.properties-Datei in Hadoop YARN.	Not available.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Ändern Sie die Docker-bezogenen Einstellungen.	Not available.
emrfs-site	Ändert die EMRFS-Einstellungen.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the



Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
		Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegionserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.	Restarts Flink history server.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.	Restarts Flink history server.
flink-log4j-session	Ändern Sie die Einstellungen von Flink log4j-session.properties für Kubernetes-/Yarn-Sitzungen.	Restarts Flink history server.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.	Restarts Flink history server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.	Not available.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Ändert die Werte in der HBase-Umgebung.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.	Not available.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.	This classification should not be reconfigured.
hdfs-env	Ändert die Werte in der HDFS-Umgebung.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei proto-hive-site .xml von HCatalog.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.	Restarts Hive WebHCat server.
hive	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Hive.	Sets configurations to launch Hive LLAP service.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.	Not available.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.	Not available.
hive-env	Ändert die Werte in der Hive-Umgebung.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.	Not available.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log	Not available.
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.	Not available.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hue-ini	Ändert die Werte in der INI-Datei in Hue.	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.	Restarts Hadoop Httpfs service.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.	Not available.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.	Restarts Hadoop-KMS service.
hadoop-kms-java-home	Das KMS-Java-Home von Hadoop ändern	Not available.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.	Not available.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Ändern der Werte in der Hudi-Umgebung.	Not available.
hudi-defaults	Ändern Sie die Werte in der hudi-defaults.conf-Datei in Hudi.	Not available.
iceberg-defaults	Ändern Sie die Werte in der iceberg-defaults.conf-Datei von Iceberg.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
delta-defaults	Ändern Sie die Werte in der delta-defaults.conf-Datei von Delta.	Not available.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.	Not available.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.	Not available.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.	Not available.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.	Not available.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.	Restarts Livy Server.
livy-env	Ändert die Werte in der Livy-Umgebung.	Restarts Livy Server.
livy-log4j2	Ändern Sie die log4j2.properties-Einstellungen für Livy.	Restarts Livy Server.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.	Restarts Hadoop MapReduce-HistoryServer.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
oozie-env	Ändert die Werte in der Oozie-Umgebung.	Restarts Oozie.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.	Restarts Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.	Not available.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.	Not available.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.	Not available.
pig-env	Ändert die Werte in der Pig-Umgebung.	Not available.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.	Restarts Oozie.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.	Not available.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)



Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-config	Ändert die Werte in der config.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.	Not available.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.	Not available.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.	Not available.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.	Not available.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.	Not available.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.	Not available.
presto-connector-lakeformation	Ändern Sie die Werte in der lakeformation.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.	Not available.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.	Not available.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.	Not available.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.	Not available.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.	Not available.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.	Not available.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.	Not available.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.	Not available.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.	Not available.
trino-log	Ändern Sie die Werte in der log.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-config	Ändern Sie die Werte in der config.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Ändern Sie die Werte in der Trino-Datei password-authenticator.properties.	Restarts Trino-Server (for Trino)
trino-env	Ändern Sie die Werte in der trino-env.sh-Datei von Trino.	Restarts Trino-Server (for Trino)
trino-node	Ändern Sie die Werte in der node.properties-Datei in Trino.	Not available.
trino-connector-blackhole	Ändern Sie die Werte in der blackhole.properties-Datei in Trino.	Not available.
trino-connector-cassandra	Ändern Sie die Werte in der cassandra.properties-Datei in Trino.	Not available.
trino-connector-delta	Ändern Sie die Werte in der delta.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-hive	Ändern Sie die Werte in der hive.properties-Datei in Trino.	Restarts Trino-Server (for Trino)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-exchange-manager	Ändern Sie die Werte in der <code>exchange-manager.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Ändern Sie die Werte in der <code>iceberg.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-hudi	Ändern Sie die Werte in der <code>hudi.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Ändern Sie die Werte in der <code>jmx.properties</code> -Datei in Trino.	Not available.
trino-connector-kafka	Ändern Sie die Werte in der <code>kafka.properties</code> -Datei in Trino.	Not available.
trino-connector-localfile	Ändern Sie die Werte in der <code>localfile.properties</code> -Datei in Trino.	Not available.
trino-connector-memory	Ändern Sie die Werte in der <code>memory.properties</code> -Datei in Trino.	Not available.
trino-connector-mongodb	Ändern Sie die Werte in der <code>mongodb.properties</code> -Datei in Trino.	Not available.
trino-connector-mysql	Ändern Sie die Werte in der <code>mysql.properties</code> -Datei in Trino.	Not available.
trino-connector-postgresql	Ändern Sie die Werte in der <code>postgresql.properties</code> -Datei in Trino.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-connector-raptor	Ändern Sie die Werte in der raptor.properties-Datei in Trino.	Not available.
trino-connector-redis	Ändern Sie die Werte in der redis.properties-Datei in Trino.	Not available.
trino-connector-redshift	Ändern Sie die Werte in der redshift.properties-Datei in Trino.	Not available.
trino-connector-tpch	Ändern Sie die Werte in der tpch.properties-Datei in Trino.	Not available.
trino-connector-tpcds	Ändern Sie die Werte in der tpcds.properties-Datei in Trino.	Not available.
ranger-kms-dbks-site	Ändert die Werte in der dbks-site.xml-Datei von Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Ändert die Werte in der Ranger KMS-Umgebung.	Restarts Ranger KMS Server.
ranger-kms-logback	Ändern Sie die Werte in der kms-logback.xml-Datei von Ranger KMS.	Not available.
ranger-kms-db-ca	Ändert die Werte für die CA-Datei auf S3 für die MySQL SSL-Verbindung mit Ranger KMS.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Ändert die Werte in der Spark-Umgebung.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.	Not available.
spark-log4j2	Ändern Sie die Werte in der log4j2.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.	Not available.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.	Not available.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.	Not available.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.	Restart Oozie and HiveServer2.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
yarn-env	Ändert die Werte in der YARN-Umgebung.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.	Restarts Zeppelin.
zeppelin-site	Ändern Sie die Konfigurationseinstellungen in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.	Restarts Zookeeper server.

## 6.12.0 Änderungsprotokoll

### Änderungsprotokoll für Version 6.12.0 und Versionshinweise

Datum	Ereignis	Beschreibung
2023-07-27	Dokumentation der Aktualisierung	Aktualisieren Sie die Java-Optionen für 6.12 und fügen Sie das Oozie-Tutorial hinzu, um JVM zu aktualisieren
2023-07-21	Die Bereitstellung ist abgeschlossen	<a href="#">Amazon EMR 6.12.0 vollständig in allen unterstützten Regionen bereitgestellt</a>
2023-07-21	Veröffentlichung von Dokumenten	Erste Veröffentlichung der Versionshinweise zu Amazon EMR 6.12.0
2023-07-12	Erstversion	Amazon EMR 6.12.0 wurde zuerst in ersten kommerziellen Regionen bereitgestellt

## Amazon-EMR-Version 6.11.1

### 6.11.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Delta](#), [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)



- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

### Informationen zur Anwendungsversion

	emr-6.11.1	emr-6.11.0	emr-6.10.1	emr-6.10.0
AWS SDK for Java	1.12.446	1.12,446	1.12,397	1,12,397
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2.12,15	2.12,15	2.12,15	2.12,15
AmazonCloudWatchAgent	-	-	-	-
Delta	2.2.0	2.2.0	2.2.0	2.2.0
Flink	1.16.0	1.16.0	1.16.0	1.16.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4.15	2.4,15	2.4,15	2.4,15
HCatalog	3.1.3	3.1.3	3.1.3	3.1.3
Hadoop	3.3.3	3.3.3	3.3.3	3.3.3
Hive	3.1.3	3.1.3	3.1.3	3.1.3
Hudi	0.13.0-amzn-0	0.13.0-amzn-0	0.12.2-amzn-0	0.12.2-amzn-0
Hue	4.11.0	4.11.0	4.10.0	4.10.0
Iceberg	1.2.0-amzn-0	1.2.0-amzn-0	1.1.0-amzn-0	1.1.0-amzn-0
JupyterEnterpriseGateway	2.6.0	2.6.0	2.6.0	2.6.0

	emr-6.11.1	emr-6.11.0	emr-6.10.1	emr-6.10.0
JupyterHub	1.4.1	1.4.1	1.5.0	1.5.0
Livy	0.7.1	0.7.1	0.7.1	0.7.1
MXNet	1.9.1	1.9.1	1.9.1	1.9.1
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	5.1.2	5.1.2	5.1.2	5.1.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0,279	0,279	0,278	0,278
Spark	3.3.2	3.3.2	3.3.1	3.3.1
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.11.0	2.11.0	2.11.0	2.11.0
Tez	0,102	0.10.2	0.10.2	0.10.2
Trino (PrestoSQL)	410	410	403	403
Zeppelin	0.10.1	0.10.1	0.10.1	0.10.1
ZooKeeper	3.5.10	3.5.10	3.5.10	3.5.10

## 6.11.1 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 6.11.1. Änderungen beziehen sich auf Version 6.11.0. Informationen zum Zeitplan für die Veröffentlichung finden Sie unter [6.11.1 Änderungsprotokoll](#).

## Änderungen, Verbesserungen und behobene Probleme

- Aufgrund von Sperrkonflikten kann ein Knoten in einen Deadlock geraten, wenn er gleichzeitig mit dem Versuch, ihn außer Betrieb zu nehmen, hinzugefügt oder entfernt wird. Infolgedessen reagiert der Hadoop Resource Manager (YARN) nicht mehr und wirkt sich auf alle eingehenden und aktuell laufenden Container aus.
- Diese Version enthält eine Änderung, die es Hochverfügbarkeitsclustern ermöglicht, nach einem Neustart den Status „Fehler“ wiederherzustellen.
- Diese Version enthält Sicherheitsupdates für Hue und HBase.
- Diese Version behebt ein Problem, bei dem Cluster, die Workloads auf Spark mit Amazon EMR ausführen, möglicherweise unbemerkt falsche Ergebnisse mit `contains`, `startsWith`, `endsWith` und `like` erhalten. Dieses Problem tritt auf, wenn Sie die Ausdrücke für partitionierte Felder verwenden, die Metadaten im Amazon EMR Hive3 Metastore Server (HMS) haben.
- Diese Version behebt ein Problem mit der Drosselung auf der Glue-Seite, wenn es keine benutzerdefinierten Funktionen (UDF) gibt.
- Diese Version behebt ein Problem, bei dem Container-Protokolle vom Knotenprotokoll-Aggregationsservice gelöscht werden, bevor Log Pusher sie im Falle einer Außerbetriebnahme von YARN an S3 weiterleiten kann.
- Diese Version behebt ein Problem mit FairShare Scheduler-Metriken, wenn Node Label für Hadoop aktiviert ist.
- Diese Version behebt ein Problem, das die Leistung von Spark beeinträchtigte, wenn Sie einen `true`-Standardwert für die `spark.yarn.heterogeneousExecutors.enabled`-Konfiguration in `spark-defaults.conf` festlegen.
- Diese Version behebt ein Problem, bei dem Reduce Task die Shuffle-Daten nicht lesen konnte. Das Problem führte zu Hive-Abfragefehlern mit einem Speicherfehler.
- Diese Version fügt dem Cluster-Skalierungs-Workflow für EMR-Cluster, auf denen Presto oder Trino ausgeführt werden, einen neuen Wiederholungsmechanismus hinzu. Diese Verbesserung verringert das Risiko, dass die Clustergrößenänderung aufgrund eines einzigen fehlgeschlagenen Größenänderungsvorgangs auf unbestimmte Zeit zum Stillstand kommt. Sie verbessert auch die Clusterauslastung, da Ihr Cluster schneller hoch- und herunterskaliert wird.
- Diese Version verbessert die Cluster-Herunterskalierungs-Logik, sodass Ihr Cluster nicht versucht, die Core-Knoten unter die für den Cluster eingestellte HDFS-Replikationsfaktor-Einstellung herunterzuskalieren. Dies entspricht Ihren Anforderungen an die Datenredundanz und verringert die Wahrscheinlichkeit, dass ein Skalierungsvorgang zum Stillstand kommt.

- Der Protokoll-Management-Daemon wurde aktualisiert, um alle Protokolle, die aktiv verwendet werden, mit offenen Datei-Handles auf dem lokalen Instance-Speicher und die zugehörigen Prozesse zu identifizieren. Dieses Upgrade stellt sicher, dass Amazon EMR die Dateien ordnungsgemäß löscht und Speicherplatz zurückgewinnt, nachdem die Protokolle in Amazon S3 archiviert wurden.
- Diese Version beinhaltet eine Erweiterung des Protokollverwaltungs-Daemons, die leere, unbenutzte Schrittverzeichnisse im lokalen Cluster-Dateisystem löscht. Eine zu große Anzahl leerer Verzeichnisse kann die Leistung der Amazon-EMR-Daemons beeinträchtigen und zu einer Überauslastung der Festplatte führen.
- Wenn Sie einen Cluster mit der neuesten Patch-Version von Amazon EMR 5.36 oder höher, 6.6 oder höher oder 7.0 oder höher starten, verwendet Amazon EMR die neueste Version von Amazon Linux 2023 oder Amazon Linux 2 für das standardmäßige Amazon EMR-AMI. Weitere Informationen finden Sie unter [Verwenden des standardmäßigen Amazon-Linux-AMI für Amazon EMR](#).

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 223.0	4,14.336	8. März 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost), China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024131.0	4,14.336	14. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024124.0	4,14.336	7. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024109.0	4,14.334	24. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 218.0	4,14.330	2. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsReleaseLabel (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202306.0	4,14.330	22. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023116.0	4,14.328	11. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023101.0	4,14.327	16. November 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023020.1	4,14.326	7. November 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023012.1	4,14.326	26. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 926.0	4,14.322	19. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)
2.0.2023 906.0	4,14.322	04. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 822.0	4,14.322	30. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.20230808,0	4,14.320	24. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 727.0	4,14.320	14. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral)

### 6.11.1 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version

sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-amzn-*EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen myapp-component der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als 2.2-amzn-2 aufgeführt.

Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1.4.2	Amazon SageMaker Spark-SDK
delta	2.2.0	Delta Lake ist ein offenes Tabellenformat für riesige analytische Datensätze
delta-standalone-connectors	0.6.0	Delta-Konnektoren bieten unterschiedliche Laufzeiten, um Delta Lake mit Engines wie Flink, Hive und Presto zu integrieren.
emr-ddb	5.1.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	3.4.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.8.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-notebook-env	1.7.0	Conda Env für EMR-Notebooks, das Jupyter Enterprise Gateway enthält

Komponente	Version	Beschreibung
emr-s3-dist-cp	2.25.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	2.4.0	EMR S3Select-Konnektor
emr-wal-cli	1.1.0	Cli wird für die emrwal list/deletion verwendet.
emrfs	2,56,0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.16.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
flink-jobmanager-config	1.16.0	Verwaltung von Ressourcen auf EMR-Knoten für Apache JobManager Flink.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.

Komponente	Version	Beschreibung
hadoop-client	3.3.3-amzn-3.1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	3.3.3-amzn-3.1	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	3.3.3-amzn-3.1	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	3.3.3-amzn-3.1	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-hdfs-journalnode	3.3.3-amzn-3.1	HDFS-Service zum Verwalten des Hadoop-Dateisystem journals auf HA-Clustern.
hadoop-httfs-server	3.3.3-amzn-3.1	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	3.3.3-amzn-3.1	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	3.3.3-amzn-3.1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	3.3.3-amzn-3.1	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.

Komponente	Version	Beschreibung
hadoop-yarn-resourcemanager	3.3.3-amzn-3.1	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	3.3.3-amzn-3.1	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	2.4.15-amzn-1.1	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	2.4.15-amzn-1.1	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	2.4.15-amzn-1.1	HBase-Befehlszeilen-Client.
hbase-rest-server	2.4.15-amzn-1.1	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	2.4.15-amzn-1.1	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hbase-operator-tools	2.4.15-amzn-1.1	Reparaturtool für Apache-HBase-Cluster.
hcatalog-client	3.1.3-amzn-4.1	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.



Komponente	Version	Beschreibung
hcatalog-server	3.1.3-amzn-4.1	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	3.1.3-amzn-4.1	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	3.1.3-amzn-4.1	Hive-Befehlszeilen-Client.
hive-hbase	3.1.3-amzn-4.1	Hive-hbase client.
hive-metastore-server	3.1.3-amzn-4.1	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	3.1.3-amzn-4.1	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.13.0-amzn-0	Inkrementelles Verarbeitungs-Framework zur Versorgung der Datenpipeline mit geringer Latenz und hoher Effizienz.
hudi-presto	0.13.0-amzn-0	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hudi-trino	0.13.0-amzn-0	Bündel-Bibliothek zum Ausführen von Trino mit Hudi.

Komponente	Version	Beschreibung
hudi-spark	0.13.0-amzn-0	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.
hue-server	4.11.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
iceberg	1.2.0-amzn-0	Apache Iceberg ist ein offenes Tabellenformat für sehr große analytische Datensätze
jupyterhub	1.4.1	Multi-User-Server für Jupyter-Notebooks
livy-server	0.7.1-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP- und Reverse-Proxy-Server.
mxnet	1.9.1	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.68+	MariaDB-Datenbankserver.
nvidia-cuda	11,8,0	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.2.1	Oozie-Befehlszeilen-Client.
oozie-server	5.2.1	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	4.5.0	Open Source Computer Vision Library.

Komponente	Version	Beschreibung
phoenix-library	5.1.2	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-connectors	5.1.2	Apache-Phoenix-Konnektoren für Spark-3
phoenix-query-server	5.1.2	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.279-amzn-0	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.279-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.279-amzn-0	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
trino-coordinator	410-amzn-0	Service zur Annahme von Abfragen und Verwaltung der Abfrageausführung der Trino-Worker.
trino-worker	410-amzn-0	Service für das Ausführen von Teilen einer Abfrage.

Komponente	Version	Beschreibung
trino-client	410-amzn-0	Trino-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Trino-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	4.0.2	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	2.0.0	Apache Ranger Key Management System
spark-client	3.3.2-amzn-0.1	Spark-Befehlszeilen-Clients.
spark-history-server	3.3.2-amzn-0.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	3.3.2-amzn-0.1	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	3.3.2-amzn-0.1	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
spark-rapids	23.02.0-amzn-0	Nvidia-Spark-RAPIDS-Plugin, das Apache Spark mit GPUs beschleunigt.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.

Komponente	Version	Beschreibung
tensorflow	2.11.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.10.2-amzn-2.1	Die Tez-YARN-Anwendung und -Bibliotheken.
tez-on-worker	0.10.2-amzn-2.1	Die Tez-YARN-Anwendung und Bibliotheken für Worker-Knoten.
webserver	2.4.41+	Apache HTTP-Server.
zeppelin-server	0.10.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.5.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.5.10	ZooKeeper Befehlszeilen-Cli-ent.

### 6.11.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Aktionen zur Neukonfiguration treten auf, wenn Sie eine Konfiguration für Instance-Gruppen in einem laufenden Cluster angeben. Amazon EMR initiiert nur Rekonfigurationsaktionen für die Klassifizierungen, die Sie ändern. Weitere Informationen finden Sie unter [Eine Instance-Gruppe in einem laufenden Cluster neu konfigurieren](#).

#### emr-6.11.1-Klassifizierungen

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
capacity-scheduler	Ändert die Werte in der capacity-scheduler.xml-Datei in Hadoop.	Restarts the ResourceManager service.
container-executor	Ändern Sie die Werte in der Datei „container-executor.cfg“ Datei von Hadoop YARN.	Not available.
container-log4j	Ändert die Werte in der container-log4j.properties-Datei in Hadoop YARN.	Not available.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Ändern Sie die Docker-bezogenen Einstellungen.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
emrfs-site	Ändert die EMRFS-Einstellungen.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegistrator, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.	Restarts Flink history server.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.	Restarts Flink history server.
flink-log4j-session	Ändern Sie die Einstellungen von Flink log4j-session.properties für Kubernetes-/Yarn-Sitzungen.	Restarts Flink history server.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.	Restarts Flink history server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.	Not available.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.	Not available.



Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Ändert die Werte in der HBase-Umgebung.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.	Not available.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.	This classification should not be reconfigured.
hdfs-env	Ändert die Werte in der HDFS-Umgebung.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei proto-hive-site .xml von HCatalog.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.	Restarts Hive WebHCat server.
hive	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Hive.	Sets configurations to launch Hive LLAP service.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.	Not available.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.	Not available.
hive-env	Ändert die Werte in der Hive-Umgebung.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.	Not available.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log	Not available.
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.	Not available.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hue-ini	Ändert die Werte in der INI-Datei in Hue.	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.	Restarts Hadoop Httpfs service.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.	Not available.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.	Not available.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Ändern der Werte in der Hudi-Umgebung.	Not available.
hudi-defaults	Ändern Sie die Werte in der hudi-defaults.conf-Datei in Hudi.	Not available.
iceberg-defaults	Ändern Sie die Werte in der iceberg-defaults.conf-Datei von Iceberg.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
delta-defaults	Ändern Sie die Werte in der delta-defaults.conf-Datei von Delta.	Not available.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.	Not available.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.	Not available.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.	Not available.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.	Not available.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.	Restarts Livy Server.
livy-env	Ändert die Werte in der Livy-Umgebung.	Restarts Livy Server.
livy-log4j2	Ändern Sie die log4j2.properties-Einstellungen für Livy.	Restarts Livy Server.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.	Restarts Hadoop MapReduce-HistoryServer.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
oozie-env	Ändert die Werte in der Oozie-Umgebung.	Restarts Oozie.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.	Restarts Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.	Not available.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.	Not available.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.	Not available.
pig-env	Ändert die Werte in der Pig-Umgebung.	Not available.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.	Restarts Oozie.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.	Not available.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-config	Ändert die Werte in der config.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.	Not available.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.	Not available.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.	Not available.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.	Not available.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.	Not available.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.	Not available.
presto-connector-lakeformation	Ändern Sie die Werte in der lakeformation.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.	Not available.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.	Not available.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.	Not available.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.	Not available.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.	Not available.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.	Not available.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.	Not available.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.	Not available.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.	Not available.



Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.	Not available.
trino-log	Ändern Sie die Werte in der log.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-config	Ändern Sie die Werte in der config.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Ändern Sie die Werte in der Trino-Datei password-authenticator.properties.	Restarts Trino-Server (for Trino)
trino-env	Ändern Sie die Werte in der trino-env.sh-Datei von Trino.	Restarts Trino-Server (for Trino)
trino-node	Ändern Sie die Werte in der node.properties-Datei in Trino.	Not available.
trino-connector-blackhole	Ändern Sie die Werte in der blackhole.properties-Datei in Trino.	Not available.
trino-connector-cassandra	Ändern Sie die Werte in der cassandra.properties-Datei in Trino.	Not available.
trino-connector-delta	Ändern Sie die Werte in der delta.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-hive	Ändern Sie die Werte in der hive.properties-Datei in Trino.	Restarts Trino-Server (for Trino)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-exchange-manager	Ändern Sie die Werte in der exchange-manager.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Ändern Sie die Werte in der iceberg.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-hudi	Ändern Sie die Werte in der hudi.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Ändern Sie die Werte in der jmx.properties-Datei in Trino.	Not available.
trino-connector-kafka	Ändern Sie die Werte in der kafka.properties-Datei in Trino.	Not available.
trino-connector-localfile	Ändern Sie die Werte in der localfile.properties-Datei in Trino.	Not available.
trino-connector-memory	Ändern Sie die Werte in der memory.properties-Datei in Trino.	Not available.
trino-connector-mongodb	Ändern Sie die Werte in der mongodb.properties-Datei in Trino.	Not available.
trino-connector-mysql	Ändern Sie die Werte in der mysql.properties-Datei in Trino.	Not available.
trino-connector-postgresql	Ändern Sie die Werte in der postgresql.properties-Datei in Trino.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-connector-raptor	Ändern Sie die Werte in der raptor.properties-Datei in Trino.	Not available.
trino-connector-redis	Ändern Sie die Werte in der redis.properties-Datei in Trino.	Not available.
trino-connector-redshift	Ändern Sie die Werte in der redshift.properties-Datei in Trino.	Not available.
trino-connector-tpch	Ändern Sie die Werte in der tpch.properties-Datei in Trino.	Not available.
trino-connector-tpcds	Ändern Sie die Werte in der tpcds.properties-Datei in Trino.	Not available.
ranger-kms-dbks-site	Ändert die Werte in der dbks-site.xml-Datei von Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Ändert die Werte in der Ranger KMS-Umgebung.	Restarts Ranger KMS Server.
ranger-kms-logback	Ändern Sie die Werte in der kms-logback.xml-Datei von Ranger KMS.	Not available.
ranger-kms-db-ca	Ändert die Werte für die CA-Datei auf S3 für die MySQL SSL-Verbindung mit Ranger KMS.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Ändert die Werte in der Spark-Umgebung.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.	Not available.
spark-log4j2	Ändern Sie die Werte in der log4j2.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.	Not available.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.	Not available.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.	Not available.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.	Restart Oozie and HiveServer2.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
yarn-env	Ändert die Werte in der YARN-Umgebung.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.	Restarts Zeppelin.
zeppelin-site	Ändern Sie die Konfigurationseinstellungen in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.	Restarts Zookeeper server.

## 6.11.1 Änderungsprotokoll

### Änderungsprotokoll für Version 6.11.1 und Versionshinweise

Datum	Ereignis	Beschreibung
2023-08-30	Versionshinweise der Aktualisierung	Den Versionshinweisen wurden mehrere Korrekturen im Zusammenhang mit der Steuerebene hinzugefügt
2023-08-21	Veröffentlichung von Dokumenten	Erste Veröffentlichung der Versionshinweise zu Amazon EMR 6.11.1
2023-08-16	Die Bereitstellung ist abgeschlossen	<a href="#">Amazon EMR 6.11.1 vollständig in allen unterstützten Regionen bereitgestellt</a>
2023-08-04	Erstversion	Amazon EMR 6.11.1 wurde zuerst in begrenzten kommerziellen Regionen bereitgestellt

## Amazon-EMR-Version 6.11.0

### 6.11.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Delta](#), [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

### Informationen zur Anwendungsversion

	emr-6.11.0	emr-6.10.1	emr-6.10.0	emr-6.9.1
AWS SDK for Java	1.12.446	1.12,397	1,12,397	1.12,170
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2.12,15	2.12,15	2.12,15	2.12,15
AmazonCloudWatchAgent	-	-	-	-
Delta	2.2.0	2.2.0	2.2.0	2.1.0
Flink	1.16.0	1.16.0	1.16.0	1.15.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4.15	2.4,15	2.4,15	2.4.13
HCatalog	3.1.3	3.1.3	3.1.3	3.1.3
Hadoop	3.3.3	3.3.3	3.3.3	3.3.3
Hive	3.1.3	3.1.3	3.1.3	3.1.3
Hudi	0.13.0-amzn-0	0.12.2-amzn-0	0.12.2-amzn-0	0.12.1-amzn-0
Hue	4.11.0	4.10.0	4.10.0	4.10.0
Iceberg	1.2.0-amzn-0	1.1.0-amzn-0	1.1.0-amzn-0	0.14.1-amzn-0

	emr-6.11.0	emr-6.10.1	emr-6.10.0	emr-6.9.1
JupyterEnterpriseGateway	2.6.0	2.6.0	2.6.0	2.6.0
JupyterHub	1.4.1	1.5.0	1.5.0	1.4.1
Livy	0.7.1	0.7.1	0.7.1	0.7.1
MXNet	1.9.1	1.9.1	1.9.1	1.9.1
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	5.1.2	5.1.2	5.1.2	5.1.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0,279	0,278	0,278	0,276
Spark	3.3.2	3.3.1	3.3.1	3.3.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.11.0	2.11.0	2.11.0	2.10.0
Tez	0,102	0.10.2	0.10.2	0.10.2
Trino (PrestoSQL)	410	403	403	398
Zeppelin	0,10,1	0.10.1	0.10.1	0.10.1
ZooKeeper	3.5.10	3.5.10	3.5.10	3.5.10

## 6.11.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 6.11.0. Änderungen beziehen sich auf Version 6.10.0. Informationen zum Zeitplan der Veröffentlichung finden Sie im [Änderungsprotokoll](#).



## Neue Features

- Amazon EMR 6.11.0 unterstützt Apache Spark 3.3.2-amzn-0, Apache Spark RAPIDS 23.02.0-amzn-0, CUDA 11.8.0, Apache Hudi 0.13.0-amzn-0, Apache Iceberg 1.2.0-amzn-0, Trino 410-amzn-0 und PrestoDB 0.279-amzn-0.

## Änderungen, Verbesserungen und behobene Probleme

- Mit Amazon EMR 6.11.0 wurde der DynamoDB-Konnektor auf Version 5.0.0 aktualisiert. Version 5.0.0 verwendet AWS SDK for Java 2.x In früheren Versionen wurde AWS SDK for Java 1.x verwendet. Aufgrund dieses Upgrades empfehlen wir Ihnen dringend, Ihren Code zu testen, bevor Sie den DynamoDB-Konnektor mit Amazon EMR 6.11 verwenden.
- Wenn der DynamoDB-Konnektor für Amazon EMR 6.11.0 den DynamoDB-Service aufruft, verwendet er den Region-Wert, den Sie für die Eigenschaft `dynamodb.endpoint` angeben. Wir empfehlen, dass Sie auch `dynamodb.region` konfigurieren, wenn Sie `dynamodb.endpoint` verwenden, und dass beide Eigenschaften auf dieselbe AWS-Region abzielen. Wenn Sie verwenden `dynamodb.endpoint` und nicht konfigurierend `dynamodb.region`, gibt der DynamoDB-Connector für Amazon EMR 6.11.0 eine ungültige Region-Ausnahme zurück und versucht, Ihre AWS-Region Informationen aus dem Amazon EC2 EC2-Instance-Metadaten-Service (IMDS) abzugleichen. Wenn der Konnektor die Region nicht aus IMDS abrufen kann, verwendet er standardmäßig USA Ost (Nord-Virginia) (`us-east-1`). Der folgende Fehler ist ein Beispiel für die ungültige Region-Ausnahme, die möglicherweise angezeigt wird, wenn Sie die `dynamodb.region` Eigenschaft nicht richtig konfigurieren: `error software.amazon.awssdk.services.dynamodb.model.DynamoDbException: Credential should be scoped to a valid region`. Weitere Informationen zu den Klassen, die vom AWS SDK for Java Upgrade auf 2.x betroffen sind, finden Sie im Commit [Upgrade AWS SDK for Java from 1.x auf 2.x \(#175\)](#) im GitHub Repository für den Amazon EMR — DynamoDB-Connector.
- Diese Version behebt ein Problem, bei dem Spaltendaten zu NULL werden, wenn Sie Delta Lake zum Speichern von Delta-Tabellendaten in Amazon S3 nach dem Spaltenumbenennungsvorgang verwenden. Weitere Informationen zu diesem experimentellen Feature in Delta Lake finden Sie unter [Vorgang zum Umbenennen von Spalten](#) im Delta-Lake-Benutzerhandbuch.
- Die Version 6.11.0 behebt ein Problem, das auftreten kann, wenn Sie einen Edge-Knoten erstellen, indem Sie einen der Primärknoten aus einem Cluster mit mehreren Primärknoten replizieren. Der replizierte Edge-Knoten kann zu Verzögerungen bei Scale-Down-Vorgängen oder zu einer hohen Arbeitsspeicherauslastung auf den Primärknoten führen. Weitere Informationen zum Erstellen


eines Edge-Knotens für die Kommunikation mit Ihrem EMR-Cluster finden Sie unter [Edge Node Creator](#) im `aws-samples` Repo unter GitHub

- Die Version 6.11.0 verbessert den Automatisierungsprozess, den Amazon EMR verwendet, um Amazon-EBS-Volumes nach einem Neustart erneut in eine Instance einzubinden.
- Die Version 6.11.0 behebt ein Problem, das zu zeitweiligen Lücken in den Hadoop-Metriken führte, die Amazon EMR auf Amazon veröffentlicht. CloudWatch
- Die Version 6.11.0 behebt ein Problem mit EMR-Clustern, bei dem ein Update der YARN-Konfigurationsdatei, die die Ausschlussliste der Knoten für den Cluster enthält, aufgrund einer Überauslastung der Festplatte unterbrochen wird. Das unvollständige Update behindert zukünftige Cluster-Herunterskalierungs-Vorgänge. Diese Version stellt sicher, dass Ihr Cluster fehlerfrei bleibt und dass die Skalierungsvorgänge wie erwartet funktionieren.
- Die Standardgröße des Root-Volumes wurde in Amazon EMR 6.10.0 und höher auf 15 GB erhöht. Frühere Versionen haben eine Standardgröße für das Root-Volume von 10 GB.
- Mit Hadoop 3.3.3 wurde eine Änderung in YARN ([YARN-9608](#)) eingeführt, die dafür sorgt, dass Knoten, auf denen Container laufen, außer Betrieb genommen werden, bis die Anwendung abgeschlossen ist. Diese Änderung stellt sicher, dass lokale Daten wie Shuffle-Daten nicht verloren gehen und Sie den Auftrag nicht erneut ausführen müssen. Dieser Ansatz kann auch zu einer Unterauslastung von Ressourcen in Clustern mit oder ohne aktivierter verwalteter Skalierung führen.

Bei den Amazon-EMR-Versionen 6.11.0 und höher sowie 6.8.1, 6.9.1 und 6.10.1 ist der Wert in `yarn-site.xml` von `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` auf `false` gesetzt, um dieses Problem zu beheben.

Der Fix behebt zwar die Probleme, die durch YARN-9608 eingeführt wurden, kann jedoch dazu führen, dass Hive-Aufträge aufgrund von Shuffle-Datenverlusten auf Clustern, für die verwaltete Skalierung aktiviert ist, fehlschlagen. Wir haben dieses Risiko in dieser Version verringert, indem wir auch `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-shuffle-data` für Hive-Workloads eingerichtet haben. Diese Konfiguration ist nur mit Amazon-EMR-Versionen 6.11.0 und höher verfügbar.

- Wenn Sie einen Cluster mit der neuesten Patch-Version von Amazon EMR 5.36 oder höher, 6.6 oder höher oder 7.0 oder höher starten, verwendet Amazon EMR die neueste Version von Amazon Linux 2023 oder Amazon Linux 2 für das standardmäßige Amazon EMR-AMI. Weitere Informationen finden Sie unter [Verwenden des standardmäßigen Amazon-Linux-AMI für Amazon EMR](#).

 Note

Diese Version erhält keine automatischen AMI-Updates mehr, da sie von einer weiteren Patch-Versionen abgelöst wurde. Die Patch-Version wird durch die Zahl nach dem zweiten Dezimalpunkt (6.8.**1**) gekennzeichnet. Um zu sehen, ob Sie die neueste Patch-Version verwenden, überprüfen Sie die verfügbaren Versionen im [Versionshandbuch](#) oder überprüfen Sie die Amazon-EMR-Versions-Dropdownliste, wenn Sie einen Cluster in der Konsole erstellen, oder verwenden Sie die [ListReleaseLabels](#)-API- oder [list-release-labels](#)-CLI-Aktion. Um aktuelle Informationen über Neuerscheinungen erhalten, abonnieren Sie den RSS-Feed auf der Seite [Was ist neu?](#).

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.20230808.0	4,14.320	24. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 727.0	4,14.320	14. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 719,0	4,14.320	02. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 628.0	4,14.318	12. Juli 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Zürich), Europa (Mailand), Europa (Spanien), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 612,0	4,14.314	23. Juni 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Zürich), Europa (Mailand), Europa (Spanien), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE)



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 504.1	4,14.313	16. Mai 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral)

### 6.11.0 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon

EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.4.2	Amazon SageMaker Spark-SDK
<code>delta</code>	2.2.0	Delta Lake ist ein offenes Tabellenformat für riesige analytische Datensätze
<code>delta-standalone-connectors</code>	0.6.0	Delta-Konnektoren bieten unterschiedliche Laufzeiten, um Delta Lake mit Engines wie Flink, Hive und Presto zu integrieren.
<code>emr-ddb</code>	5.1.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	3.4.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.8.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-notebook-env	1.7.0	Conda Env für EMR-Notebooks, das Jupyter Enterprise Gateway enthält
emr-s3-dist-cp	2.25.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	2.4.0	EMR S3Select-Konnektor
emr-wal-cli	1.1.0	Cli wird für die emrwal list/deletion verwendet.
emrfs	2,56,0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.16.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
flink-jobmanager-config	1.16.0	Verwaltung von Ressourcen auf EMR-Knoten für Apache JobManager Flink.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.

Komponente	Version	Beschreibung
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	3.3.3-amzn-3	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	3.3.3-amzn-3	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	3.3.3-amzn-3	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	3.3.3-amzn-3	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-hdfs-journalnode	3.3.3-amzn-3	HDFS-Service zum Verwalten des Hadoop-Dateisystem journals auf HA-Clustern.
hadoop-httpfs-server	3.3.3-amzn-3	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	3.3.3-amzn-3	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	3.3.3-amzn-3	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.

Komponente	Version	Beschreibung
hadoop-yarn-nodemanager	3.3.3-amzn-3	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	3.3.3-amzn-3	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	3.3.3-amzn-3	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	2.4.15-amzn-1	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	2.4.15-amzn-1	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	2.4.15-amzn-1	HBase-Befehlszeilen-Client.
hbase-rest-server	2.4.15-amzn-1	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	2.4.15-amzn-1	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hbase-operator-tools	2.4.15-amzn-1	Reparaturtool für Apache-HBase-Cluster.

Komponente	Version	Beschreibung
hcatalog-client	3.1.3-amzn-4	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	3.1.3-amzn-4	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	3.1.3-amzn-4	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	3.1.3-amzn-4	Hive-Befehlszeilen-Client.
hive-hbase	3.1.3-amzn-4	Hive-hbase client.
hive-metastore-server	3.1.3-amzn-4	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	3.1.3-amzn-4	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.13.0-amzn-0	Inkrementelles Verarbeitungs-Framework zur Versorgung der Datenpipeline mit geringer Latenz und hoher Effizienz.
hudi-presto	0.13.0-amzn-0	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.

Komponente	Version	Beschreibung
hudi-trino	0.13.0-amzn-0	Bündel-Bibliothek zum Ausführen von Trino mit Hudi.
hudi-spark	0.13.0-amzn-0	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.
hue-server	4.11.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
iceberg	1.2.0-amzn-0	Apache Iceberg ist ein offenes Tabellenformat für sehr große analytische Datensätze
jupyterhub	1.4.1	Multi-User-Server für Jupyter-Notebooks
livy-server	0.7.1-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP- und Reverse-Proxy-Server.
mxnet	1.9.1	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.68+	MariaDB-Datenbankserver.
nvidia-cuda	11,8,0	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.2.1	Oozie-Befehlszeilen-Client.
oozie-server	5.2.1	Service für die Annahme von Oozie Workflow-Anforderungen.

Komponente	Version	Beschreibung
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	5.1.2	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-connectors	5.1.2	Apache-Phoenix-Konnektoren für Spark-3
phoenix-query-server	5.1.2	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.279-amzn-0	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.279-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.279-amzn-0	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
trino-coordinator	410-amzn-0	Service zur Annahme von Abfragen und Verwaltung der Abfrageausführung der Trino-Worker.
trino-worker	410-amzn-0	Service für das Ausführen von Teilen einer Abfrage.



Komponente	Version	Beschreibung
trino-client	410-amzn-0	Trino-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Trino-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	4.0.2	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	2.0.0	Apache Ranger Key Management System
spark-client	3.3.2-amzn-0	Spark-Befehlszeilen-Clients.
spark-history-server	3.3.2-amzn-0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	3.3.2-amzn-0	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	3.3.2-amzn-0	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
spark-rapids	23.02.0-amzn-0	Nvidia-Spark-RAPIDS-Plugin, das Apache Spark mit GPUs beschleunigt.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.

Komponente	Version	Beschreibung
tensorflow	2.11.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.10.2-amzn-2	Die Tez-YARN-Anwendung und -Bibliotheken.
tez-on-worker	0.10.2-amzn-2	Die Tez-YARN-Anwendung und Bibliotheken für Worker-Knoten.
webserver	2.4.41+	Apache HTTP-Server.
zeppelin-server	0.10.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.5.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.5.10	ZooKeeper Befehlszeilen-Cli-ent.

## 6.11.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Aktionen zur Neukonfiguration treten auf, wenn Sie eine Konfiguration für Instance-Gruppen in einem laufenden Cluster angeben. Amazon EMR initiiert nur Rekonfigurationsaktionen für die Klassifizierungen, die Sie ändern. Weitere Informationen finden Sie unter [Eine Instance-Gruppe in einem laufenden Cluster neu konfigurieren](#).

#### emr-6.11.0-Klassifizierungen

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
capacity-scheduler	Ändert die Werte in der capacity-scheduler.xml-Datei in Hadoop.	Restarts the ResourceManager service.
container-executor	Ändern Sie die Werte in der Datei „container-executor.cfg“ Datei von Hadoop YARN.	Not available.
container-log4j	Ändert die Werte in der container-log4j.properties-Datei in Hadoop YARN.	Not available.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Ändern Sie die Docker-bezogenen Einstellungen.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
emrfs-site	Ändert die EMRFS-Einstellungen.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegistrator, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.	Restarts Flink history server.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.	Restarts Flink history server.
flink-log4j-session	Ändern Sie die Einstellungen von Flink log4j-session.properties für Kubernetes-/Yarn-Sitzungen.	Restarts Flink history server.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.	Restarts Flink history server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.	Not available.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Ändert die Werte in der HBase-Umgebung.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.	Not available.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.	This classification should not be reconfigured.
hdfs-env	Ändert die Werte in der HDFS-Umgebung.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei proto-hive-site .xml von HCatalog.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.	Restarts Hive WebHCat server.
hive	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Hive.	Sets configurations to launch Hive LLAP service.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.	Not available.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.	Not available.
hive-env	Ändert die Werte in der Hive-Umgebung.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.	Not available.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log	Not available.
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.	Not available.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.	Not available.



Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hue-ini	Ändert die Werte in der INI-Datei in Hue.	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.	Restarts Hadoop Httpfs service.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.	Not available.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.	Not available.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Ändern der Werte in der Hudi-Umgebung.	Not available.
hudi-defaults	Ändern Sie die Werte in der hudi-defaults.conf-Datei in Hudi.	Not available.
iceberg-defaults	Ändern Sie die Werte in der iceberg-defaults.conf-Datei von Iceberg.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
delta-defaults	Ändern Sie die Werte in der delta-defaults.conf-Datei von Delta.	Not available.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.	Not available.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.	Not available.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.	Not available.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.	Not available.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.	Restarts Livy Server.
livy-env	Ändert die Werte in der Livy-Umgebung.	Restarts Livy Server.
livy-log4j2	Ändern Sie die log4j2.properties-Einstellungen für Livy.	Restarts Livy Server.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.	Restarts Hadoop MapReduce-HistoryServer.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
oozie-env	Ändert die Werte in der Oozie-Umgebung.	Restarts Oozie.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.	Restarts Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.	Not available.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.	Not available.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.	Not available.
pig-env	Ändert die Werte in der Pig-Umgebung.	Not available.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.	Restarts Oozie.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.	Not available.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-config	Ändert die Werte in der config.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.	Not available.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.	Not available.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.	Not available.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.	Not available.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.	Not available.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.	Not available.
presto-connector-lakeformation	Ändern Sie die Werte in der lakeformation.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.	Not available.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.	Not available.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.	Not available.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.	Not available.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.	Not available.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.	Not available.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.	Not available.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.	Not available.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.	Not available.
trino-log	Ändern Sie die Werte in der log.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-config	Ändern Sie die Werte in der config.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Ändern Sie die Werte in der Trino-Datei password-authenticator.properties.	Restarts Trino-Server (for Trino)
trino-env	Ändern Sie die Werte in der trino-env.sh-Datei von Trino.	Restarts Trino-Server (for Trino)
trino-node	Ändern Sie die Werte in der node.properties-Datei in Trino.	Not available.
trino-connector-blackhole	Ändern Sie die Werte in der blackhole.properties-Datei in Trino.	Not available.
trino-connector-cassandra	Ändern Sie die Werte in der cassandra.properties-Datei in Trino.	Not available.
trino-connector-delta	Ändern Sie die Werte in der delta.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-hive	Ändern Sie die Werte in der hive.properties-Datei in Trino.	Restarts Trino-Server (for Trino)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-exchange-manager	Ändern Sie die Werte in der exchange-manager.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Ändern Sie die Werte in der iceberg.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-hudi	Ändern Sie die Werte in der hudi.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Ändern Sie die Werte in der jmx.properties-Datei in Trino.	Not available.
trino-connector-kafka	Ändern Sie die Werte in der kafka.properties-Datei in Trino.	Not available.
trino-connector-localfile	Ändern Sie die Werte in der localfile.properties-Datei in Trino.	Not available.
trino-connector-memory	Ändern Sie die Werte in der memory.properties-Datei in Trino.	Not available.
trino-connector-mongodb	Ändern Sie die Werte in der mongodb.properties-Datei in Trino.	Not available.
trino-connector-mysql	Ändern Sie die Werte in der mysql.properties-Datei in Trino.	Not available.
trino-connector-postgresql	Ändern Sie die Werte in der postgresql.properties-Datei in Trino.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-connector-raptor	Ändern Sie die Werte in der raptor.properties-Datei in Trino.	Not available.
trino-connector-redis	Ändern Sie die Werte in der redis.properties-Datei in Trino.	Not available.
trino-connector-redshift	Ändern Sie die Werte in der redshift.properties-Datei in Trino.	Not available.
trino-connector-tpch	Ändern Sie die Werte in der tpch.properties-Datei in Trino.	Not available.
trino-connector-tpcds	Ändern Sie die Werte in der tpcds.properties-Datei in Trino.	Not available.
ranger-kms-dbks-site	Ändert die Werte in der dbks-site.xml-Datei von Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Ändert die Werte in der Ranger KMS-Umgebung.	Restarts Ranger KMS Server.
ranger-kms-logback	Ändern Sie die Werte in der kms-logback.xml-Datei von Ranger KMS.	Not available.
ranger-kms-db-ca	Ändert die Werte für die CA-Datei auf S3 für die MySQL SSL-Verbindung mit Ranger KMS.	Not available.



Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Ändert die Werte in der Spark-Umgebung.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.	Not available.
spark-log4j2	Ändern Sie die Werte in der log4j2.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.	Not available.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.	Not available.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.	Not available.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.	Restart Oozie and HiveServer2.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
yarn-env	Ändert die Werte in der YARN-Umgebung.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.	Restarts Zeppelin.
zeppelin-site	Ändern Sie die Konfigurationseinstellungen in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.	Restarts Zookeeper server.

## 6.11.0 Änderungsprotokoll

### Änderungsprotokoll für Version 6.11.0 und Versionshinweise

Datum	Ereignis	Beschreibung
2023-08-21	Aktualisierung	Das mit Hadoop 3.3.3 eingeführte Problem wurde behoben.
2023-07-26	Aktualisierung	Neue Betriebssystem-Release-Labels 2.0.20230612.0 und 2.0.20230628.0 .
2023-06-09	Die Bereitstellung ist abgeschlossen	<a href="#">Amazon EMR 6.11.0 vollständig in allen unterstützten Regionen bereitgestellt</a>
2023-06-09	Veröffentlichung von Dokumenten	Erste Veröffentlichung der Versionshinweise zu Amazon EMR 6.11.0
2023-06-08	Erstversion	Amazon EMR 6.11.0 wurde zuerst in ersten kommerziellen Regionen bereitgestellt

## Amazon-EMR-Version 6.10.1

### 6.10.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Delta](#), [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

### Informationen zur Anwendungsversion

	emr-6.10.1	emr-6.10.0	emr-6.9.1	emr-6.9.0
AWS SDK for Java	1.12.397	1,12,397	1.12,170	1.12,170
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2.12,15	2.12,15	2.12,15	2.12,15
AmazonCloudWatchAgent	-	-	-	-
Delta	2.2.0	2.2.0	2.1.0	2.1.0
Flink	1.16.0	1.16.0	1.15.2	1.15.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4.15	2.4,15	2.4.13	2.4.13
HCatalog	3.1.3	3.1.3	3.1.3	3.1.3
Hadoop	3.3.3	3.3.3	3.3.3	3.3.3
Hive	3.1.3	3.1.3	3.1.3	3.1.3
Hudi	0.12.2-amzn-0	0.12.2-amzn-0	0.12.1-amzn-0	0.12.1-amzn-0
Hue	4.10.0	4.10.0	4.10.0	4.10.0

	emr-6.10.1	emr-6.10.0	emr-6.9.1	emr-6.9.0
Iceberg	1.1.0-amzn-0	1.1.0-amzn-0	0.14.1-amzn-0	0.14.1-amzn-0
JupyterEnterpriseGateway	2.6.0	2.6.0	2.6.0	2.6.0
JupyterHub	1.5.0	1.5.0	1.4.1	1.4.1
Livy	0.7.1	0.7.1	0.7.1	0.7.1
MXNet	1.9.1	1.9.1	1.9.1	1.9.1
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	5.1.2	5.1.2	5.1.2	5.1.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0,278	0,278	0,276	0,276
Spark	3.3.1	3.3.1	3.3.0	3.3.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.11.0	2.11.0	2.10.0	2.10.0
Tez	0,102	0.10.2	0.10.2	0.10.2
Trino (PrestoSQL)	403	403	398	398
Zeppelin	0,10,1	0.10.1	0.10.1	0.10.1
ZooKeeper	3.5.10	3.5.10	3.5.10	3.5.10

## 6.10.1 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 6.10.1. Änderungen beziehen sich auf Version 6.10.0. Informationen zum Zeitplan für die Veröffentlichung finden Sie unter [6.10.1 Änderungsprotokoll](#).

### Änderungen, Verbesserungen und behobene Probleme

- Aufgrund von Sperrkonflikten kann ein Knoten in einen Deadlock geraten, wenn er gleichzeitig mit dem Versuch, ihn außer Betrieb zu nehmen, hinzugefügt oder entfernt wird. Infolgedessen reagiert der Hadoop Resource Manager (YARN) nicht mehr und wirkt sich auf alle eingehenden und aktuell laufenden Container aus.
- Mit Hadoop 3.3.3 wurde eine Änderung in YARN ([YARN-9608](#)) eingeführt, die dafür sorgt, dass Knoten, auf denen Container laufen, außer Betrieb genommen werden, bis die Anwendung abgeschlossen ist. Diese Änderung stellt sicher, dass lokale Daten wie Shuffle-Daten nicht verloren gehen und Sie den Auftrag nicht erneut ausführen müssen. Dieser Ansatz kann auch zu einer Unterauslastung von Ressourcen in Clustern mit oder ohne aktivierter verwalteter Skalierung führen.

Bei den Amazon-EMR-Versionen 6.11.0 und höher sowie 6.8.1, 6.9.1 und 6.10.1 ist der Wert in `yarn-site.xml` von `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` auf `false` gesetzt, um dieses Problem zu beheben.

Der Fix behebt zwar die Probleme, die durch YARN-9608 eingeführt wurden, kann jedoch dazu führen, dass Hive-Aufträge aufgrund von Shuffle-Datenverlusten auf Clustern, für die verwaltete Skalierung aktiviert ist, fehlschlagen. Wir haben dieses Risiko in dieser Version verringert, indem wir auch `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-shuffle-data` für Hive-Workloads eingerichtet haben. Diese Konfiguration ist nur mit Amazon-EMR-Versionen 6.11.0 und höher verfügbar.

- Der Metrics Collector sendet nach einem Failover des Primärknotens in Clustern mit der Instance-Gruppenkonfiguration keine Metriken an die Steuerebene.
- Diese Version enthält eine Änderung, die es Hochverfügbarkeitsclustern ermöglicht, nach einem Neustart den Status „Fehler“ wiederherzustellen.
- Diese Version enthält Sicherheitsupdates für Hue und HBase.
- Diese Version behebt ein Problem, bei dem Cluster, die Workloads auf Spark mit Amazon EMR ausführen, möglicherweise unbemerkt falsche Ergebnisse mit `contains`, `startsWith`,

endsWith und like erhalten. Dieses Problem tritt auf, wenn Sie die Ausdrücke für partitionierte Felder verwenden, die Metadaten im Amazon EMR Hive3 Metastore Server (HMS) haben.

- Diese Version behebt ein Problem mit der Drosselung auf der Glue-Seite, wenn es keine benutzerdefinierten Funktionen (UDF) gibt.
- Diese Version behebt ein Problem, bei dem Container-Protokolle vom Knotenprotokoll-Aggregationservice gelöscht werden, bevor Log Pusher sie im Falle einer Außerbetriebnahme von YARN an S3 weiterleiten kann.
- Diese Version behebt ein Problem mit FairShare Scheduler-Metriken, wenn Node Label für Hadoop aktiviert ist.
- Diese Version behebt ein Problem, das die Leistung von Spark beeinträchtigte, wenn Sie einen true-Standardwert für die spark.yarn.heterogeneousExecutors.enabled-Konfiguration in spark-defaults.conf festlegen.
- Diese Version behebt ein Problem, bei dem Reduce Task die Shuffle-Daten nicht lesen konnte. Das Problem führte zu Hive-Abfragefehlern mit einem Speicherfehler.
- Diese Version fügt dem Cluster-Skalierungs-Workflow für EMR-Cluster, auf denen Presto oder Trino ausgeführt werden, einen neuen Wiederholungsmechanismus hinzu. Diese Verbesserung verringert das Risiko, dass die Clustergrößenänderung aufgrund eines einzigen fehlgeschlagenen Größenänderungsvorgangs auf unbestimmte Zeit zum Stillstand kommt. Sie verbessert auch die Clusterauslastung, da Ihr Cluster schneller hoch- und herunterskaliert wird.
- Diese Version verbessert die Cluster-Herunterskalierungs-Logik, sodass Ihr Cluster nicht versucht, die Core-Knoten unter die für den Cluster eingestellte HDFS-Replikationsfaktor-Einstellung herunterzuskalieren. Dies entspricht Ihren Anforderungen an die Datenredundanz und verringert die Wahrscheinlichkeit, dass ein Skalierungsvorgang zum Stillstand kommt.
- Der Protokoll-Management-Daemon wurde aktualisiert, um alle Protokolle, die aktiv verwendet werden, mit offenen Datei-Handles auf dem lokalen Instance-Speicher und die zugehörigen Prozesse zu identifizieren. Dieses Upgrade stellt sicher, dass Amazon EMR die Dateien ordnungsgemäß löscht und Speicherplatz zurückgewinnt, nachdem die Protokolle in Amazon S3 archiviert wurden.
- Diese Version beinhaltet eine Erweiterung des Protokollverwaltungs-Daemons, die leere, unbenutzte Schrittverzeichnisse im lokalen Cluster-Dateisystem löscht. Eine zu große Anzahl leerer Verzeichnisse kann die Leistung der Amazon-EMR-Daemons beeinträchtigen und zu einer Überauslastung der Festplatte führen.
- Diese Version behebt ein Problem, das auftreten kann, wenn Sie einen Edge-Knoten erstellen, indem Sie einen der Primärknoten aus einem Cluster mit mehreren Primärknoten replizieren. Der

replizierte Edge-Knoten kann zu Verzögerungen bei Scale-Down-Vorgängen oder zu einer hohen Arbeitsspeicherauslastung auf den Primärknoten führen. Weitere Informationen zum Erstellen eines Edge-Knotens für die Kommunikation mit Ihrem EMR-Cluster finden Sie unter [Edge Node Creator](#) im `aws-samples` Repo unter [GitHub](#)

- Diese Version verbessert den Automatisierungsprozess, den Amazon EMR verwendet, um Amazon-EBS-Volumes nach einem Neustart erneut in eine Instance einzubinden.
- Diese Version behebt ein Problem, das zu zeitweiligen Lücken in den Hadoop-Metriken führte, die Amazon EMR auf Amazon veröffentlicht. CloudWatch
- Diese Version behebt ein Problem mit EMR-Clustern, bei dem ein Update der YARN-Konfigurationsdatei, die die Ausschlussliste der Knoten für den Cluster enthält, aufgrund einer Überauslastung der Festplatte unterbrochen wird. Das unvollständige Update behindert zukünftige Cluster-Herunterskalierungs-Vorgänge. Diese Version stellt sicher, dass Ihr Cluster fehlerfrei bleibt und dass die Skalierungsvorgänge wie erwartet funktionieren.
- Wenn Sie einen Cluster mit der neuesten Patch-Version von Amazon EMR 5.36 oder höher, 6.6 oder höher oder 7.0 oder höher starten, verwendet Amazon EMR die neueste Version von Amazon Linux 2023 oder Amazon Linux 2 für das standardmäßige Amazon EMR-AMI. Weitere Informationen finden Sie unter [Verwenden des standardmäßigen Amazon-Linux-AMI für Amazon EMR](#).

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 223.0	4,14.336	8. März 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost), China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024131.0	4,14.336	14. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024124.0	4,14.336	7. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsReleaseLabel (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024109.0	4,14.334	24. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 218.0	4,14.330	2. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202306.0	4,14.330	22. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 116.0	4,14.328	11. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023101.0	4,14.327	16. November 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023020.1	4,14.326	7. November 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023012.1	4,14.326	26. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsReleaseLabel (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 926.0	4,14.322	19. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)
2.0.2023 906.0	4,14.322	04. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 822.0	4,14.322	30. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.20230808,0	4,14.320	24. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 727.0	4,14.320	14. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral)

### 6.10.1 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version



sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-amzn-*EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen myapp-component der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als 2.2-amzn-2 aufgeführt.

Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1.4.2	Amazon SageMaker Spark-SDK
delta	2.2.0	Delta Lake ist ein offenes Tabellenformat für riesige analytische Datensätze
emr-ddb	4.16.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	3.3.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.7.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-notebook-env	1.7.0	Conda Env für EMR-Notebooks, das Jupyter Enterprise Gateway enthält
emr-s3-dist-cp	2.24.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	2.3.0	EMR S3Select-Konnektor

Komponente	Version	Beschreibung
emr-wal-cli	1.0.0	Cli wird für die emrwal list/deletion verwendet.
emrfs	2,55,0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.16.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
flink-jobmanager-config	1.16.0	Verwaltung von Ressourcen auf EMR-Knoten für Apache JobManager Flink.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	3.3.3-amzn-2.1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".

Komponente	Version	Beschreibung
hadoop-hdfs-datanode	3.3.3-amzn-2.1	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	3.3.3-amzn-2.1	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	3.3.3-amzn-2.1	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-hdfs-journalnode	3.3.3-amzn-2.1	HDFS-Service zum Verwalten des Hadoop-Dateisystem journals auf HA-Clustern.
hadoop-https-server	3.3.3-amzn-2.1	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	3.3.3-amzn-2.1	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	3.3.3-amzn-2.1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	3.3.3-amzn-2.1	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	3.3.3-amzn-2.1	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.

Komponente	Version	Beschreibung
hadoop-yarn-timeline-server	3.3.3-amzn-2.1	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	2.4.15-amzn-0.1	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	2.4.15-amzn-0.1	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	2.4.15-amzn-0.1	HBase-Befehlszeilen-Client.
hbase-rest-server	2.4.15-amzn-0.1	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	2.4.15-amzn-0.1	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hbase-operator-tools	2.4.15-amzn-0.1	Reparaturtool für Apache-HBase-Cluster.
hcatalog-client	3.1.3-amzn-3.1	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	3.1.3-amzn-3.1	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).

Komponente	Version	Beschreibung
hcatalog-webhcat-server	3.1.3-amzn-3.1	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	3.1.3-amzn-3.1	Hive-Befehlszeilen-Client.
hive-hbase	3.1.3-amzn-3.1	Hive-hbase client.
hive-metastore-server	3.1.3-amzn-3.1	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	3.1.3-amzn-3.1	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.12.2-amzn-0	Inkrementelles Verarbeitungs-Framework zur Versorgung der Datenpipeline mit geringer Latenz und hoher Effizienz.
hudi-presto	0.12.2-amzn-0	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hudi-trino	0.12.2-amzn-0	Bündel-Bibliothek zum Ausführen von Trino mit Hudi.
hudi-spark	0.12.2-amzn-0	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.
hue-server	4.10.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.

Komponente	Version	Beschreibung
iceberg	1.1.0-amzn-0	Apache Iceberg ist ein offenes Tabellenformat für sehr große analytische Datensätze
jupyterhub	1.5.0	Multi-User-Server für Jupyter-Notebooks
livy-server	0.7.1-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP- und Reverse-Proxy-Server.
mxnet	1.9.1	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.68+	MariaDB-Datenbankserver.
nvidia-cuda	11,8,0	Nvidia-Treiber und Cuda-Toolkit
oozie-client	5.2.1	Oozie-Befehlszeilen-Client.
oozie-server	5.2.1	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	5.1.2	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-connectors	6.0.0-SNAPSHOT	Apache-Phoenix-Konnektoren für Spark-3

Komponente	Version	Beschreibung
phoenix-query-server	6.0.0	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.278.1-amzn-0	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.278.1-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.278.1-amzn-0	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
trino-coordinator	403-amzn-0	Service zur Annahme von Abfragen und Verwaltung der Abfrageausführung der Trino-Worker.
trino-worker	403-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
trino-client	403-amzn-0	Trino-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Trino-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.

Komponente	Version	Beschreibung
r	4.0.2	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	2.0.0	Apache Ranger Key Management System
spark-client	3.3.1-amzn-0.1	Spark-Befehlszeilen-Clients.
spark-history-server	3.3.1-amzn-0.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	3.3.1-amzn-0.1	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	3.3.1-amzn-0.1	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
spark-rapids	22.12.0-amzn-0	Nvidia-Spark-RAPIDS-Plugin, das Apache Spark mit GPUs beschleunigt.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	2.11.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.10.2-amzn-1.1	Die Tez-YARN-Anwendung und -Bibliotheken.



Komponente	Version	Beschreibung
tez-on-worker	0.10.2-amzn-1.1	Die Tez-YARN-Anwendung und Bibliotheken für Worker-Knoten.
webserver	2.4.41+	Apache HTTP-Server.
zeppelin-server	0.10.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.5.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.5.10	ZooKeeper Befehlszeilen-Cli-ent.

## 6.10.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Aktionen zur Neukonfiguration treten auf, wenn Sie eine Konfiguration für Instance-Gruppen in einem laufenden Cluster angeben. Amazon EMR initiiert nur Rekonfigurationsaktionen für die Klassifizierungen, die Sie ändern. Weitere Informationen finden Sie unter [Eine Instance-Gruppe in einem laufenden Cluster neu konfigurieren](#).

## emr-6.10.1-Klassifizierungen

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
capacity-scheduler	Ändert die Werte in der capacity-scheduler.xml-Datei in Hadoop.	Restarts the ResourceManager service.
container-executor	Ändern Sie die Werte in der Datei „container-executor.cfg“ Datei von Hadoop YARN.	Not available.
container-log4j	Ändert die Werte in der container-log4j.properties-Datei in Hadoop YARN.	Not available.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Ändern Sie die Docker-bezogenen Einstellungen.	Not available.
emrfs-site	Ändert die EMRFS-Einstellungen.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
		Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegionserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.	Restarts Flink history server.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.	Restarts Flink history server.
flink-log4j-session	Ändern Sie die Einstellungen von Flink log4j-session.properties für Kubernetes-/Yarn-Sitzungen.	Restarts Flink history server.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.	Restarts Flink history server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.	Not available.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Ändert die Werte in der HBase-Umgebung.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.	Not available.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.	This classification should not be reconfigured.
hdfs-env	Ändert die Werte in der HDFS-Umgebung.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei proto-hive-site .xml von HCatalog.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.	Restarts Hive WebHCat server.
hive	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Hive.	Sets configurations to launch Hive LLAP service.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.	Not available.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.	Not available.
hive-env	Ändert die Werte in der Hive-Umgebung.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.	Not available.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log	Not available.
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.	Not available.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hue-ini	Ändert die Werte in der INI-Datei in Hue.	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.	Restarts Hadoop Httpfs service.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.	Not available.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.	Not available.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Ändern der Werte in der Hudi-Umgebung.	Not available.
hudi-defaults	Ändern Sie die Werte in der hudi-defaults.conf-Datei in Hudi.	Not available.
iceberg-defaults	Ändern Sie die Werte in der iceberg-defaults.conf-Datei von Iceberg.	Not available.



Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
delta-defaults	Ändern Sie die Werte in der delta-defaults.conf-Datei von Delta.	Not available.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.	Not available.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.	Not available.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.	Not available.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.	Not available.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.	Restarts Livy Server.
livy-env	Ändert die Werte in der Livy-Umgebung.	Restarts Livy Server.
livy-log4j2	Ändern Sie die log4j2.properties-Einstellungen für Livy.	Restarts Livy Server.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.	Restarts Hadoop MapReduce-HistoryServer.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
oozie-env	Ändert die Werte in der Oozie-Umgebung.	Restarts Oozie.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.	Restarts Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.	Not available.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.	Not available.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.	Not available.
pig-env	Ändert die Werte in der Pig-Umgebung.	Not available.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.	Restarts Oozie.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.	Not available.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-config	Ändert die Werte in der config.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.	Not available.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.	Not available.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.	Not available.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.	Not available.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.	Not available.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.	Not available.
presto-connector-lakeformation	Ändern Sie die Werte in der lakeformation.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.	Not available.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.	Not available.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.	Not available.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.	Not available.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.	Not available.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.	Not available.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.	Not available.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.	Not available.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.	Not available.
trino-log	Ändern Sie die Werte in der log.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-config	Ändern Sie die Werte in der config.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Ändern Sie die Werte in der Trino-Datei password-authenticator.properties.	Restarts Trino-Server (for Trino)
trino-env	Ändern Sie die Werte in der trino-env.sh-Datei von Trino.	Restarts Trino-Server (for Trino)
trino-node	Ändern Sie die Werte in der node.properties-Datei in Trino.	Not available.
trino-connector-blackhole	Ändern Sie die Werte in der blackhole.properties-Datei in Trino.	Not available.
trino-connector-cassandra	Ändern Sie die Werte in der cassandra.properties-Datei in Trino.	Not available.
trino-connector-delta	Ändern Sie die Werte in der delta.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-hive	Ändern Sie die Werte in der hive.properties-Datei in Trino.	Restarts Trino-Server (for Trino)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-exchange-manager	Ändern Sie die Werte in der exchange-manager.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Ändern Sie die Werte in der iceberg.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-hudi	Ändern Sie die Werte in der hudi.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Ändern Sie die Werte in der jmx.properties-Datei in Trino.	Not available.
trino-connector-kafka	Ändern Sie die Werte in der kafka.properties-Datei in Trino.	Not available.
trino-connector-localfile	Ändern Sie die Werte in der localfile.properties-Datei in Trino.	Not available.
trino-connector-memory	Ändern Sie die Werte in der memory.properties-Datei in Trino.	Not available.
trino-connector-mongodb	Ändern Sie die Werte in der mongodb.properties-Datei in Trino.	Not available.
trino-connector-mysql	Ändern Sie die Werte in der mysql.properties-Datei in Trino.	Not available.
trino-connector-postgresql	Ändern Sie die Werte in der postgresql.properties-Datei in Trino.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-connector-raptor	Ändern Sie die Werte in der raptor.properties-Datei in Trino.	Not available.
trino-connector-redis	Ändern Sie die Werte in der redis.properties-Datei in Trino.	Not available.
trino-connector-redshift	Ändern Sie die Werte in der redshift.properties-Datei in Trino.	Not available.
trino-connector-tpch	Ändern Sie die Werte in der tpch.properties-Datei in Trino.	Not available.
trino-connector-tpcds	Ändern Sie die Werte in der tpcds.properties-Datei in Trino.	Not available.
ranger-kms-dbks-site	Ändert die Werte in der dbks-site.xml-Datei von Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Ändert die Werte in der Ranger KMS-Umgebung.	Restarts Ranger KMS Server.
ranger-kms-logback	Ändern Sie die Werte in der kms-logback.xml-Datei von Ranger KMS.	Not available.
ranger-kms-db-ca	Ändert die Werte für die CA-Datei auf S3 für die MySQL SSL-Verbindung mit Ranger KMS.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Ändert die Werte in der Spark-Umgebung.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.	Not available.
spark-log4j2	Ändern Sie die Werte in der log4j2.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.	Not available.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.	Not available.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.	Not available.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.	Restart Oozie and HiveServer2.



Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
yarn-env	Ändert die Werte in der YARN-Umgebung.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.	Restarts Zeppelin.
zeppelin-site	Ändern Sie die Konfigurationseinstellungen in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.	Restarts Zookeeper server.

## 6.10.1 Änderungsprotokoll

### Änderungsprotokoll für Version 6.10.1 und Versionshinweise

Datum	Ereignis	Beschreibung
2023-08-30	Versionshinweise der Aktualisierung	Den Versionshinweisen wurden mehrere Korrekturen im Zusammenhang mit der Steuerebene hinzugefügt
2023-08-21	Veröffentlichung von Dokumenten	Erste Veröffentlichung der Versionshinweise zu Amazon EMR 6.10.1
2023-08-16	Die Bereitstellung ist abgeschlossen	<a href="#">Amazon EMR 6.10.1 vollständig in allen unterstützten Regionen bereitgestellt</a>
2023-08-04	Erstversion	Amazon EMR 6.10.1 wurde zuerst in begrenzten kommerziellen Regionen bereitgestellt

## Amazon-EMR-Version 6.10.0

### 6.10.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Delta](#), [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

### Informationen zur Anwendungsversion

	emr-6.10.0	emr-6.9.1	emr-6.9.0	emr-6.8.1
AWS SDK for Java	1.12.397	1.12,170	1.12,170	1.12,170
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2.12,15	2.12,15	2.12,15	2.12,15
AmazonCloudWatchAgent	-	-	-	-
Delta	2.2.0	2.1.0	2.1.0	-
Flink	1.16.0	1.15.2	1.15.2	1.15.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4.15	2.4.13	2.4.13	2.4.12
HCatalog	3.1.3	3.1.3	3.1.3	3.1.3
Hadoop	3.3.3	3.3.3	3.3.3	3.2.1
Hive	3.1.3	3.1.3	3.1.3	3.1.3
Hudi	0.12.2-amzn-0	0.12.1-amzn-0	0.12.1-amzn-0	0.11.1-amzn-0
Hue	4.10.0	4.10.0	4.10.0	4.10.0
Iceberg	1.1.0-amzn-0	0.14.1-amzn-0	0.14.1-amzn-0	0.14.0-amzn-0

	emr-6.10.0	emr-6.9.1	emr-6.9.0	emr-6.8.1
JupyterEnterpriseGateway	2.6.0	2.6.0	2.6.0	2.1.0
JupyterHub	1.5.0	1.4.1	1.4.1	1.4.1
Livy	0.7.1	0.7.1	0.7.1	0.7.1
MXNet	1.9.1	1.9.1	1.9.1	1.9.1
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	5.1.2	5.1.2	5.1.2	5.1.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0,278	0,276	0,276	0,273
Spark	3.3.1	3.3.0	3.3.0	3.3.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.11.0	2.10.0	2.10.0	2,9,1
Tez	0.10.2	0.10.2	0.10.2	0.9.2
Trino (PrestoSQL)	403	398	398	388
Zeppelin	0,10,1	0.10.1	0.10.1	0.10.1
ZooKeeper	3.5.10	3.5.10	3.5.10	3.5.10

## 6.10.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 6.10.0. Änderungen beziehen sich auf Version 6.9.0. Informationen zum Zeitplan der Veröffentlichung finden Sie im [Änderungsprotokoll](#).

## Neue Features

- Amazon EMR 6.10.0 unterstützt Apache Spark 3.3.1, Apache Spark RAPIDS 22.12.0, CUDA 11.8.0, Apache Hudi 0.12.2-amzn-0, Apache Iceberg 1.1.0-amzn-0, Trino 403 und PrestoDB 0.278.1.
- Amazon EMR 6.10.0 enthält einen nativen Trino-Hudi-Konnektor, der Lesezugriff auf Daten in Hudi-Tabellen ermöglicht. Sie können den Konnektor mit `trino-cli --catalog hudi` aktivieren und den Konnektor für Ihre Anforderungen mit `trino-connector-hudi` konfigurieren. Durch die native Integration mit Amazon EMR müssen Sie `trino-connector-hive` nicht mehr zum Abfragen von Hudi-Tabellen verwenden. Eine Liste der unterstützten Konfigurationen mit dem neuen Konnektor finden Sie auf der [Hudi-Konnektor-Seite](#) der Trino-Dokumentation.
- Amazon-EMR-Versionen 6.10.0 und höher unterstützen die Apache-Zeppelin-Integration mit Apache Flink. Weitere Informationen finden Sie unter [Arbeiten mit Flink-Aufträgen von Zeppelin in Amazon EMR](#).

## Bekannte Probleme

- Mit Hadoop 3.3.3 wurde eine Änderung in YARN ([YARN-9608](#)) eingeführt, die dafür sorgt, dass Knoten, auf denen Container laufen, außer Betrieb genommen werden, bis die Anwendung abgeschlossen ist. Diese Änderung stellt sicher, dass lokale Daten wie Shuffle-Daten nicht verloren gehen und Sie den Auftrag nicht erneut ausführen müssen. Dieser Ansatz kann auch zu einer Unterauslastung von Ressourcen in Clustern mit oder ohne aktivierter verwalteter Skalierung führen.

Um dieses Problem in Amazon EMR 6.10.0 zu umgehen, können Sie den Wert `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` auf `false` in `yarn-site.xml` setzen. In den Amazon-EMR-Versionen 6.11.0 und höher sowie 6.8.1, 6.9.1 und 6.10.1 ist die Konfiguration standardmäßig auf `false` eingestellt, um dieses Problem zu beheben.

## Änderungen, Verbesserungen und behobene Probleme

- Amazon EMR 6.10.0 entfernt die Abhängigkeit von der `minimal-json.jar` für die [Amazon-Redshift-Integration für Apache Spark](#) und fügt automatisch die erforderlichen Spark-Redshift-bezogenen JAR-Dateien zum Executor-Klassenpfad für Spark hinzu: `spark-redshift.jar`, `spark-avro.jar` und `RedshiftJDBC.jar`.

- Die Version 6.10.0 verbessert den On-Cluster-Log-Management-Daemon zur Überwachung zusätzlicher Protokollordner in Ihrem EMR-Cluster. Durch diese Verbesserung werden Szenarien mit übermäßiger Festplattenauslastung minimiert.
- Die Version 6.10.0 startet den Protokoll-Management-Daemon auf dem Cluster automatisch neu, wenn er beendet wird. Durch diese Verbesserung wird das Risiko verringert, dass Knoten aufgrund übermäßiger Festplattenauslastung als fehlerhaft erscheinen.
- Amazon EMR 6.10.0 unterstützt regionale Endpunkte für die EMRFS-Benutzerzuordnung.
- Die Standardgröße des Root-Volumes wurde in Amazon EMR 6.10.0 und höher auf 15 GB erhöht. Frühere Versionen haben eine Standardgröße für das Root-Volume von 10 GB.
- Die Version 6.10.0 behebt ein Problem, das dazu führte, dass Spark-Aufträge zum Stillstand kamen, wenn sich alle verbleibenden Spark-Ausführern auf einem außer Betrieb genommenen Host mit dem YARN-Ressourcenmanager befanden.
- Bei Amazon EMR 6.6.0 bis 6.9.x haben INSERT-Abfragen mit dynamischer Partition und einer ORDER-BY- oder SORT-BY-Klausel immer zwei Reduzierungen. Dieses Problem wird durch die OSS-Änderung [HIVE-20703](#) verursacht, die die Optimierung dynamischer Sortierpartitionen einer kostenbasierten Entscheidung unterstellt. Wenn Ihr Workload keine Sortierung dynamischer Partitionen erfordert, empfehlen wir Ihnen, die `hive.optimize.sort.dynamic.partition.threshold`-Eigenschaft auf -1 zu setzen, um das neue Feature zu deaktivieren und die korrekt berechnete Anzahl von Reduzierern zu erhalten. Dieses Problem wurde behoben in OSS Hive als Teil von [HIVE-22269](#) und in Amazon EMR 6.10.0 behoben.
- Wenn Sie einen Cluster mit der neuesten Patch-Version von Amazon EMR 5.36 oder höher, 6.6 oder höher oder 7.0 oder höher starten, verwendet Amazon EMR die neueste Version von Amazon Linux 2023 oder Amazon Linux 2 für das standardmäßige Amazon EMR-AMI. Weitere Informationen finden Sie unter [Verwenden des standardmäßigen Amazon-Linux-AMI für Amazon EMR](#).

#### Note

Diese Version erhält keine automatischen AMI-Updates mehr, da sie von einer weiteren Patch-Versionen abgelöst wurde. Die Patch-Version wird durch die Zahl nach dem zweiten Dezimalpunkt (6.8.**1**) gekennzeichnet. Um zu sehen, ob Sie die neueste Patch-Version verwenden, überprüfen Sie die verfügbaren Versionen im [Versionshandbuch](#) oder überprüfen Sie die Amazon-EMR-Versions-Dropdownliste, wenn Sie einen Cluster in der Konsole erstellen, oder verwenden Sie die [ListReleaseLabels](#)-API- oder [list-](#)

[release-labels](#)-CLI-Aktion. Um aktuelle Informationen über Neuerscheinungen erhalten, abonnieren Sie den RSS-Feed auf der Seite [Was ist neu?](#).

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 808.0	4,14.320	24. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 727.0	4,14.320	14. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv)



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 719,0	4,14.320	02. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 628.0	4,14.318	12. Juli 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Zürich), Europa (Mailand), Europa (Spanien), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 612,0	4,14.314	23. Juni 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Zürich), Europa (Mailand), Europa (Spanien), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 504.1	4,14.313	16. Mai 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Zürich), Europa (Mailand), Europa (Spanien), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 418,0	4,14.311	3. Mai 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Zürich), Europa (Mailand), Europa (Spanien), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 404.1	4,14.311	18. April 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE)
2.0.2023 404.0	4,14.311	10. April 2023	USA Ost (Nord-Virginia), Europa (Paris)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 320.0	4,14.309	30. März 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202307,0	4,14.304	22. Februar 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE)

## 6.10.0 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version



sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-amzn-*EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen myapp-component der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als 2.2-amzn-2 aufgeführt.

Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1.4.2	Amazon SageMaker Spark-SDK
delta	2.2.0	Delta Lake ist ein offenes Tabellenformat für riesige analytische Datensätze
emr-ddb	4.16.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	3.3.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.7.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-notebook-env	1.7.0	Conda Env für EMR-Notebooks, das Jupyter Enterprise Gateway enthält
emr-s3-dist-cp	2.24.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	2.3.0	EMR S3Select-Konnektor

Komponente	Version	Beschreibung
emr-wal-cli	1.0.0	Cli wird für die emrwal list/deletion verwendet.
emrfs	2,55,0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.16.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
flink-jobmanager-config	1.16.0	Verwaltung von Ressourcen auf EMR-Knoten für Apache JobManager Flink.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	3.3.3-amzn-2	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".

Komponente	Version	Beschreibung
hadoop-hdfs-datanode	3.3.3-amzn-2	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	3.3.3-amzn-2	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	3.3.3-amzn-2	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-hdfs-journalnode	3.3.3-amzn-2	HDFS-Service zum Verwalten des Hadoop-Dateisystem journals auf HA-Clustern.
hadoop-https-server	3.3.3-amzn-2	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	3.3.3-amzn-2	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	3.3.3-amzn-2	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	3.3.3-amzn-2	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	3.3.3-amzn-2	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.

Komponente	Version	Beschreibung
hadoop-yarn-timeline-server	3.3.3-amzn-2	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	2.4.15-amzn-0	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	2.4.15-amzn-0	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	2.4.15-amzn-0	HBase-Befehlszeilen-Client.
hbase-rest-server	2.4.15-amzn-0	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	2.4.15-amzn-0	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hbase-operator-tools	2.4.15-amzn-0	Reparaturtool für Apache-HBase-Cluster.
hcatalog-client	3.1.3-amzn-3	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	3.1.3-amzn-3	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).

Komponente	Version	Beschreibung
hcatalog-webhcat-server	3.1.3-amzn-3	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	3.1.3-amzn-3	Hive-Befehlszeilen-Client.
hive-hbase	3.1.3-amzn-3	Hive-hbase client.
hive-metastore-server	3.1.3-amzn-3	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	3.1.3-amzn-3	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.12.2-amzn-0	Inkrementelles Verarbeitungs-Framework zur Versorgung der Datenpipeline mit geringer Latenz und hoher Effizienz.
hudi-presto	0.12.2-amzn-0	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hudi-trino	0.12.2-amzn-0	Bündel-Bibliothek zum Ausführen von Trino mit Hudi.
hudi-spark	0.12.2-amzn-0	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.
hue-server	4.10.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.

Komponente	Version	Beschreibung
iceberg	1.1.0-amzn-0	Apache Iceberg ist ein offenes Tabellenformat für sehr große analytische Datensätze
jupyterhub	1.5.0	Multi-User-Server für Jupyter-Notebooks
livy-server	0.7.1-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP- und Reverse-Proxy-Server.
mxnet	1.9.1	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.68+	MariaDB-Datenbankserver.
nvidia-cuda	11,8,0	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.2.1	Oozie-Befehlszeilen-Client.
oozie-server	5.2.1	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	5.1.2	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-connectors	6.0.0-SNAPSHOT	Apache-Phoenix-Konnektoren für Spark-3

Komponente	Version	Beschreibung
phoenix-query-server	6.0.0	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.278.1-amzn-0	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.278.1-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.278.1-amzn-0	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
trino-coordinator	403-amzn-0	Service zur Annahme von Abfragen und Verwaltung der Abfrageausführung der Trino-Worker.
trino-worker	403-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
trino-client	403-amzn-0	Trino-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Trino-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.

Komponente	Version	Beschreibung
r	4.0.2	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	2.0.0	Apache Ranger Key Management System
spark-client	3.3.1-amzn-0	Spark-Befehlszeilen-Clients.
spark-history-server	3.3.1-amzn-0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	3.3.1-amzn-0	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	3.3.1-amzn-0	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
spark-rapids	22.12.0-amzn-0	Nvidia-Spark-RAPIDS-Plugin, das Apache Spark mit GPUs beschleunigt.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	2.11.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.10.2-amzn-1	Die Tez-YARN-Anwendung und -Bibliotheken.



Komponente	Version	Beschreibung
tez-on-worker	0.10.2-amzn-1	Die Tez-YARN-Anwendung und Bibliotheken für Worker-Knoten.
webserver	2.4.41+	Apache HTTP-Server.
zeppelin-server	0.10.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.5.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.5.10	ZooKeeper Befehlszeilen-Cli-ent.

## 6.10.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Aktionen zur Neukonfiguration treten auf, wenn Sie eine Konfiguration für Instance-Gruppen in einem laufenden Cluster angeben. Amazon EMR initiiert nur Rekonfigurationsaktionen für die Klassifizierungen, die Sie ändern. Weitere Informationen finden Sie unter [Eine Instance-Gruppe in einem laufenden Cluster neu konfigurieren](#).

## emr-6.10.0-Klassifizierungen

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
capacity-scheduler	Ändert die Werte in der capacity-scheduler.xml-Datei in Hadoop.	Restarts the ResourceManager service.
container-executor	Ändern Sie die Werte in der Datei „container-executor.cfg“ Datei von Hadoop YARN.	Not available.
container-log4j	Ändert die Werte in der container-log4j.properties-Datei in Hadoop YARN.	Not available.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Ändern Sie die Docker-bezogenen Einstellungen.	Not available.
emrfs-site	Ändert die EMRFS-Einstellungen.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
		Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegistrator, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.	Restarts Flink history server.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.	Restarts Flink history server.
flink-log4j-session	Ändern Sie die Einstellungen von Flink log4j-session.properties für Kubernetes-/Yarn-Sitzungen.	Restarts Flink history server.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.	Restarts Flink history server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.	Not available.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Ändert die Werte in der HBase-Umgebung.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.	Not available.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.	This classification should not be reconfigured.
hdfs-env	Ändert die Werte in der HDFS-Umgebung.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei proto-hive-site .xml von HCatalog.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.	Restarts Hive WebHCat server.
hive	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Hive.	Sets configurations to launch Hive LLAP service.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.	Not available.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.	Not available.
hive-env	Ändert die Werte in der Hive-Umgebung.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.	Not available.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log	Not available.
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.	Not available.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hue-ini	Ändert die Werte in der INI-Datei in Hue.	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.	Restarts Hadoop Httpfs service.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.	Not available.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.	Not available.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Ändern der Werte in der Hudi-Umgebung.	Not available.
hudi-defaults	Ändern Sie die Werte in der hudi-defaults.conf-Datei in Hudi.	Not available.
iceberg-defaults	Ändern Sie die Werte in der iceberg-defaults.conf-Datei von Iceberg.	Not available.



Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
delta-defaults	Ändern Sie die Werte in der delta-defaults.conf-Datei von Delta.	Not available.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.	Not available.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.	Not available.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.	Not available.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.	Not available.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.	Restarts Livy Server.
livy-env	Ändert die Werte in der Livy-Umgebung.	Restarts Livy Server.
livy-log4j2	Ändern Sie die log4j2.properties-Einstellungen für Livy.	Restarts Livy Server.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.	Restarts Hadoop MapReduce-HistoryServer.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
oozie-env	Ändert die Werte in der Oozie-Umgebung.	Restarts Oozie.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.	Restarts Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.	Not available.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.	Not available.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.	Not available.
pig-env	Ändert die Werte in der Pig-Umgebung.	Not available.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.	Restarts Oozie.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.	Not available.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-config	Ändert die Werte in der config.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.	Not available.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.	Not available.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.	Not available.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.	Not available.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.	Not available.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.	Not available.
presto-connector-lakeformation	Ändern Sie die Werte in der lakeformation.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.	Not available.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.	Not available.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.	Not available.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.	Not available.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.	Not available.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.	Not available.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.	Not available.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.	Not available.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.	Not available.
trino-log	Ändern Sie die Werte in der log.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-config	Ändern Sie die Werte in der config.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Ändern Sie die Werte in der Trino-Datei password-authenticator.properties.	Restarts Trino-Server (for Trino)
trino-env	Ändern Sie die Werte in der trino-env.sh-Datei von Trino.	Restarts Trino-Server (for Trino)
trino-node	Ändern Sie die Werte in der node.properties-Datei in Trino.	Not available.
trino-connector-blackhole	Ändern Sie die Werte in der blackhole.properties-Datei in Trino.	Not available.
trino-connector-cassandra	Ändern Sie die Werte in der cassandra.properties-Datei in Trino.	Not available.
trino-connector-delta	Ändern Sie die Werte in der delta.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-hive	Ändern Sie die Werte in der hive.properties-Datei in Trino.	Restarts Trino-Server (for Trino)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-exchange-manager	Ändern Sie die Werte in der exchange-manager.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Ändern Sie die Werte in der iceberg.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-hudi	Ändern Sie die Werte in der hudi.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Ändern Sie die Werte in der jmx.properties-Datei in Trino.	Not available.
trino-connector-kafka	Ändern Sie die Werte in der kafka.properties-Datei in Trino.	Not available.
trino-connector-localfile	Ändern Sie die Werte in der localfile.properties-Datei in Trino.	Not available.
trino-connector-memory	Ändern Sie die Werte in der memory.properties-Datei in Trino.	Not available.
trino-connector-mongodb	Ändern Sie die Werte in der mongodb.properties-Datei in Trino.	Not available.
trino-connector-mysql	Ändern Sie die Werte in der mysql.properties-Datei in Trino.	Not available.
trino-connector-postgresql	Ändern Sie die Werte in der postgresql.properties-Datei in Trino.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-connector-raptor	Ändern Sie die Werte in der raptor.properties-Datei in Trino.	Not available.
trino-connector-redis	Ändern Sie die Werte in der redis.properties-Datei in Trino.	Not available.
trino-connector-redshift	Ändern Sie die Werte in der redshift.properties-Datei in Trino.	Not available.
trino-connector-tpch	Ändern Sie die Werte in der tpch.properties-Datei in Trino.	Not available.
trino-connector-tpcds	Ändern Sie die Werte in der tpcds.properties-Datei in Trino.	Not available.
ranger-kms-dbks-site	Ändert die Werte in der dbks-site.xml-Datei von Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Ändert die Werte in der Ranger KMS-Umgebung.	Restarts Ranger KMS Server.
ranger-kms-logback	Ändern Sie die Werte in der kms-logback.xml-Datei von Ranger KMS.	Not available.
ranger-kms-db-ca	Ändert die Werte für die CA-Datei auf S3 für die MySQL SSL-Verbindung mit Ranger KMS.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Ändert die Werte in der Spark-Umgebung.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.	Not available.
spark-log4j2	Ändern Sie die Werte in der log4j2.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.	Not available.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.	Not available.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.	Not available.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.	Restart Oozie and HiveServer2.



Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
yarn-env	Ändert die Werte in der YARN-Umgebung.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.	Restarts Zeppelin.
zeppelin-site	Ändern Sie die Konfigurationseinstellungen in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.	Restarts Zookeeper server.

## 6.10.0 Änderungsprotokoll

### Änderungsprotokoll für Version 6.10.0 und Versionshinweise

Datum	Ereignis	Beschreibung
2023-08-21	Aktualisierung	Es wurde ein bekanntes Problem hinzugefügt, das mit Hadoop 3.3.3 eingefügt wurde.
2023-07-26	Aktualisierung	Neue Betriebssystem-Release-Labels 2.0.20230612.0 und 2.0.20230628.0 .
2023-03-02	Die Bereitstellung ist abgeschlossen	<a href="#">Amazon EMR 6.10 vollständig in allen unterstützten Regionen bereitgestellt</a>
2023-03-02	Veröffentlichung von Dokumenten	Erste Veröffentlichung der Versionshinweise zu Amazon EMR 6.10
2023-02-27	Erstversion	Amazon EMR 6.10 wurde zuerst in begrenzten kommerziellen Regionen bereitgestellt

## Amazon-EMR-Version 6.9.1

### 6.9.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Delta](#), [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

### Informationen zur Anwendungsversion

	emr-6.9.1	emr-6.9.0	emr-6.8.1	emr-6.8.0
AWS SDK for Java	1.12.170	1.12,170	1.12,170	1.12,170
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2.12,15	2.12,15	2.12,15	2.12,15
AmazonCloudWatchAgent	-	-	-	-
Delta	2.1.0	2.1.0	-	-
Flink	1.15.2	1.15.2	1.15.1	1.15.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4.13	2.4.13	2.4.12	2.4.12
HCatalog	3.1.3	3.1.3	3.1.3	3.1.3
Hadoop	3.3.3	3.3.3	3.2.1	3.2.1
Hive	3.1.3	3.1.3	3.1.3	3.1.3
Hudi	0.12.1-amzn-0	0.12.1-amzn-0	0.11.1-amzn-0	0.11.1-amzn-0
Hue	4.10.0	4.10.0	4.10.0	4.10.0

	emr-6.9.1	emr-6.9.0	emr-6.8.1	emr-6.8.0
Iceberg	0.14.1-amzn-0	0.14.1-amzn-0	0.14.0-amzn-0	0.14.0-amzn-0
JupyterEnterpriseGateway	2.6.0	2.6.0	2.1.0	2.1.0
JupyterHub	1.4.1	1.4.1	1.4.1	1.4.1
Livy	0.7.1	0.7.1	0.7.1	0.7.1
MXNet	1.9.1	1.9.1	1.9.1	1.9.1
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	5.1.2	5.1.2	5.1.2	5.1.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0,276	0,276	0,273	0,273
Spark	3.3.0	3.3.0	3.3.0	3.3.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.10.0	2.10.0	2,9,1	2.9.1
Tez	0.10.2	0.10.2	0.9.2	0.9.2
Trino (PrestoSQL)	398	398	388	388
Zeppelin	0,10,1	0.10.1	0.10.1	0.10.1
ZooKeeper	3.5.10	3.5.10	3.5.10	3.5.10

## 6.9.1 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 6.9.1. Änderungen beziehen sich auf Version 6.9.0. Informationen zum Zeitplan für die Veröffentlichung finden Sie unter [6.9.1 Änderungsprotokoll](#).

### Änderungen, Verbesserungen und behobene Probleme

- Mit Hadoop 3.3.3 wurde eine Änderung in YARN ([YARN-9608](#)) eingeführt, die dafür sorgt, dass Knoten, auf denen Container laufen, außer Betrieb genommen werden, bis die Anwendung abgeschlossen ist. Diese Änderung stellt sicher, dass lokale Daten wie Shuffle-Daten nicht verloren gehen und Sie den Auftrag nicht erneut ausführen müssen. Dieser Ansatz kann auch zu einer Unterauslastung von Ressourcen in Clustern mit oder ohne aktivierter verwalteter Skalierung führen.

Bei den Amazon-EMR-Versionen 6.11.0 und höher sowie 6.8.1, 6.9.1 und 6.10.1 ist der Wert in `yarn-site.xml` von `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` auf `false` gesetzt, um dieses Problem zu beheben.

Der Fix behebt zwar die Probleme, die durch YARN-9608 eingeführt wurden, kann jedoch dazu führen, dass Hive-Aufträge aufgrund von Shuffle-Datenverlusten auf Clustern, für die verwaltete Skalierung aktiviert ist, fehlschlagen. Wir haben dieses Risiko in dieser Version verringert, indem wir auch `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-shuffle-data` für Hive-Workloads eingerichtet haben. Diese Konfiguration ist nur mit Amazon-EMR-Versionen 6.11.0 und höher verfügbar.

- Der Metrics Collector sendet nach einem Failover des Primärknotens in Clustern mit der Instance-Gruppenkonfiguration keine Metriken an die Steuerebene.
- Diese Version macht Wiederholungsversuche bei fehlgeschlagenen HTTP-Anfragen an die Endpunkte der Metrics Collector überflüssig.
- Diese Version enthält eine Änderung, die es Hochverfügbarkeitsclustern ermöglicht, nach einem Neustart den Status „Fehler“ wiederherzustellen.
- Diese Version behebt ein Problem, bei dem große, von Benutzern erstellte UUIDs zu Überlaufausnahmen führten.
- Diese Version behebt Timeout-Probleme beim Amazon-EMR-Rekonfigurationsprozess.
- Diese Version enthält Sicherheitskorrekturen.
- Diese Version behebt ein Problem, bei dem Cluster, die Workloads auf Spark mit Amazon EMR ausführen, möglicherweise unbemerkt falsche Ergebnisse mit `contains`, `startsWith`,

endsWith und like erhalten. Dieses Problem tritt auf, wenn Sie die Ausdrücke für partitionierte Felder verwenden, die Metadaten im Amazon EMR Hive3 Metastore Server (HMS) haben.

- Bei Amazon EMR 6.6.0 bis 6.9.x haben INSERT-Abfragen mit dynamischer Partition und einer ORDER-BY- oder SORT-BY-Klausel immer zwei Reduzierungen. Dieses Problem wird durch die OSS-Änderung [HIVE-20703](#) verursacht, die die Optimierung dynamischer Sortierpartitionen einer kostenbasierten Entscheidung unterstellt. Wenn Ihr Workload keine Sortierung dynamischer Partitionen erfordert, empfehlen wir Ihnen, die `hive.optimize.sort.dynamic.partition.threshold`-Eigenschaft auf -1 zu setzen, um das neue Feature zu deaktivieren und die korrekt berechnete Anzahl von Reduzierern zu erhalten. Dieses Problem wurde behoben in OSS Hive als Teil von [HIVE-22269](#) und in Amazon EMR 6.10.0 behoben.
- In Hive kann es zu Datenverlusten kommen, wenn Sie HDFS als Scratch-Verzeichnis verwenden und die Option „Kleine Dateien zusammenführen“ aktiviert haben und die Tabelle statische Partitionspfade enthält.
- Diese Version behebt ein Leistungsproblem mit Hive, wenn das Zusammenführen kleiner Dateien (standardmäßig deaktiviert) am Ende des ETL-Auftrags aktiviert ist.
- Diese Version behebt ein Problem mit der Drosselung auf der Glue-Seite, wenn es keine benutzerdefinierten Funktionen (UDF) gibt.
- Diese Version behebt ein Problem, bei dem Container-Protokolle vom Knotenprotokoll-Aggregationsservice gelöscht werden, bevor Log Pusher sie im Falle einer Außerbetriebnahme von YARN an S3 weiterleiten kann.
- Diese Version behebt den Umgang mit komprimierten/archivierten Dateien mit persistenter Speicherdateiverfolgung für HBase.
- Diese Version behebt ein Problem, das die Leistung von Spark beeinträchtigte, wenn Sie einen `true`-Standardwert für die `spark.yarn.heterogeneousExecutors.enabled`-Konfiguration in `spark-defaults.conf` festlegen.
- Diese Version behebt ein Problem, bei dem Reduce Task die Shuffle-Daten nicht lesen konnte. Das Problem führte zu Hive-Abfragefehlern mit einem Speicherfehler.
- Diese Version behebt ein Problem, das dazu führte, dass der Node Provisioner fehlschlug, wenn der HDFS NameNode (NN) -Dienst während des Knotenaustauschs im Safemode hängen blieb.
- Diese Version fügt dem Cluster-Skalierungs-Workflow für EMR-Cluster, auf denen Presto oder Trino ausgeführt werden, einen neuen Wiederholungsmechanismus hinzu. Diese Verbesserung verringert das Risiko, dass die Clustergrößenänderung aufgrund eines einzigen fehlgeschlagenen

Größenänderungsvorgangs auf unbestimmte Zeit zum Stillstand kommt. Sie verbessert auch die Clusterauslastung, da Ihr Cluster schneller hoch- und herunterskaliert wird.

- Diese Version verbessert die Cluster-Herunterskalierungs-Logik, sodass Ihr Cluster nicht versucht, die Core-Knoten unter die für den Cluster eingestellte HDFS-Replikationsfaktor-Einstellung herunterzuskalieren. Dies entspricht Ihren Anforderungen an die Datenredundanz und verringert die Wahrscheinlichkeit, dass ein Skalierungsvorgang zum Stillstand kommt.
- Der Protokoll-Management-Daemon wurde aktualisiert, um alle Protokolle, die aktiv verwendet werden, mit offenen Datei-Handles auf dem lokalen Instance-Speicher und die zugehörigen Prozesse zu identifizieren. Dieses Upgrade stellt sicher, dass Amazon EMR die Dateien ordnungsgemäß löscht und Speicherplatz zurückgewinnt, nachdem die Protokolle in Amazon S3 archiviert wurden.
- Diese Version beinhaltet eine Erweiterung des Protokollverwaltungs-Daemons, die leere, unbenutzte Schrittverzeichnisse im lokalen Cluster-Dateisystem löscht. Eine zu große Anzahl leerer Verzeichnisse kann die Leistung der Amazon-EMR-Daemons beeinträchtigen und zu einer Überauslastung der Festplatte führen.
- Diese Version behebt ein Problem, das auftreten kann, wenn Sie einen Edge-Knoten erstellen, indem Sie einen der Primärknoten aus einem Cluster mit mehreren Primärknoten replizieren. Der replizierte Edge-Knoten kann zu Verzögerungen bei Scale-Down-Vorgängen oder zu einer hohen Arbeitsspeicherauslastung auf den Primärknoten führen. Weitere Informationen zum Erstellen eines Edge-Knotens für die Kommunikation mit Ihrem EMR-Cluster finden Sie unter [Edge Node Creator](#) im `aws-samples` Repo unter GitHub
- Diese Version verbessert den Automatisierungsprozess, den Amazon EMR verwendet, um Amazon-EBS-Volumes nach einem Neustart erneut in eine Instance einzubinden.
- Diese Version behebt ein Problem, das zu zeitweiligen Lücken in den Hadoop-Metriken führte, die Amazon EMR auf Amazon veröffentlicht. CloudWatch
- Diese Version behebt ein Problem mit EMR-Clustern, bei dem ein Update der YARN-Konfigurationsdatei, die die Ausschlussliste der Knoten für den Cluster enthält, aufgrund einer Überauslastung der Festplatte unterbrochen wird. Das unvollständige Update behindert zukünftige Cluster-Herunterskalierungs-Vorgänge. Diese Version stellt sicher, dass Ihr Cluster fehlerfrei bleibt und dass die Skalierungsvorgänge wie erwartet funktionieren.
- Diese Version verbessert den On-Cluster-Log-Management-Daemon zur Überwachung zusätzlicher Protokollordner in Ihrem EMR-Cluster. Durch diese Verbesserung werden Szenarien mit übermäßiger Festplattenauslastung minimiert.

- Diese Version startet den Protokoll-Management-Daemon auf dem Cluster automatisch neu, wenn er beendet wird. Durch diese Verbesserung wird das Risiko verringert, dass Knoten aufgrund übermäßiger Festplattenauslastung als fehlerhaft erscheinen.
- Wenn Sie einen Cluster mit der neuesten Patch-Version von Amazon EMR 5.36 oder höher, 6.6 oder höher oder 7.0 oder höher starten, verwendet Amazon EMR die neueste Version von Amazon Linux 2023 oder Amazon Linux 2 für das standardmäßige Amazon EMR-AMI. Weitere Informationen finden Sie unter [Verwenden des standardmäßigen Amazon-Linux-AMI für Amazon EMR](#).

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 223.0	4,14.336	8. März 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerik



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			a (São Paulo), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost), China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024131.0	4,14.336	14. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost), China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 124.0	4,14.336	7. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost), China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024109.0	4,14.334	24. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost), China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 218.0	4,14.330	2. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202306,0	4,14.330	22. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Kanada (Zentral), Israel AWS GovCloud (Tel Aviv), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 116.0	4,14.328	11. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Kanada (Zentral), Israel AWS GovCloud (Tel Aviv), AWS GovCloud (US-West),



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023101.0	4,14.327	16. November 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Kanada (Zentral), Israel AWS GovCloud (Tel Aviv), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023020.1	4,14.326	7. November 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Kanada (Zentral), Israel AWS GovCloud (Tel Aviv), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023012.1	4,14.326	26. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Kanada (Zentral), Israel AWS GovCloud (Tel Aviv), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 926.0	4,14.322	19. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Kanada (Zentral), Israel AWS GovCloud (Tel Aviv), AWS GovCloud (US-West),



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)
2.0.2023 906,0	4,14.322	04. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 822.0	4,14.322	30. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.20230808,0	4,14.320	24. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 727.0	4,14.320	14. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral)

### 6.9.1 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version

sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-amzn-*EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen myapp-component der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als 2.2-amzn-2 aufgeführt.

Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1.4.2	Amazon SageMaker Spark-SDK
delta	2.1.0	Delta Lake ist ein offenes Tabellenformat für riesige analytische Datensätze
emr-ddb	4.16.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	3.3.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.6.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-notebook-env	1.7.0	Conda Env für EMR-Notebooks, das Jupyter Enterprise Gateway enthält
emr-s3-dist-cp	2.23.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	2.2.0	EMR S3Select-Konnektor

Komponente	Version	Beschreibung
emrfs	2,54,0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.15.2	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
flink-jobmanager-config	1.15.2	Verwaltung von Ressourcen auf EMR-Knoten für Apache JobManager Flink.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	3.3.3-amzn-1.1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	3.3.3-amzn-1.1	HDFS-Service auf Knotenebene zum Speichern von Blöcken.

Komponente	Version	Beschreibung
hadoop-hdfs-library	3.3.3-amzn-1.1	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	3.3.3-amzn-1.1	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-hdfs-journalnode	3.3.3-amzn-1.1	HDFS-Service zum Verwalten des Hadoop-Dateisystem journals auf HA-Clustern.
hadoop-httpfs-server	3.3.3-amzn-1.1	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	3.3.3-amzn-1.1	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	3.3.3-amzn-1.1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	3.3.3-amzn-1.1	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	3.3.3-amzn-1.1	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.

Komponente	Version	Beschreibung
hadoop-yarn-timeline-server	3.3.3-amzn-1.1	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	2.4.13-amzn-0.1	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	2.4.13-amzn-0.1	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	2.4.13-amzn-0.1	HBase-Befehlszeilen-Client.
hbase-rest-server	2.4.13-amzn-0.1	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	2.4.13-amzn-0.1	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hbase-operator-tools	2.4.13-amzn-0.1	Reparaturtool für Apache-HBase-Cluster.
hcatalog-client	3.1.3-amzn-2.1	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	3.1.3-amzn-2.1	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).



Komponente	Version	Beschreibung
hcatalog-webhcat-server	3.1.3-amzn-2.1	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	3.1.3-amzn-2.1	Hive-Befehlszeilen-Client.
hive-hbase	3.1.3-amzn-2.1	Hive-hbase client.
hive-metastore-server	3.1.3-amzn-2.1	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	3.1.3-amzn-2.1	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.12.1-amzn-0	Inkrementelles Verarbeitungs-Framework zur Versorgung der Datenpipeline mit geringer Latenz und hoher Effizienz.
hudi-presto	0.12.1-amzn-0	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hudi-trino	0.12.1-amzn-0	Bündel-Bibliothek zum Ausführen von Trino mit Hudi.
hudi-spark	0.12.1-amzn-0	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.
hue-server	4.10.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.

Komponente	Version	Beschreibung
iceberg	0.14.1-amzn-0	Apache Iceberg ist ein offenes Tabellenformat für sehr große analytische Datensätze
jupyterhub	1.4.1	Multi-User-Server für Jupyter-Notebooks
livy-server	0.7.1-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP- und Reverse-Proxy-Server.
mxnet	1.9.1	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.68+	MariaDB-Datenbankserver.
nvidia-cuda	11,7,0	Nvidia-Treiber und Cuda-Toolkit
oozie-client	5.2.1	Oozie-Befehlszeilen-Client.
oozie-server	5.2.1	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	5.1.2	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-connectors	6.0.0-SNAPSHOT	Apache-Phoenix-Konnektoren für Spark-3

Komponente	Version	Beschreibung
phoenix-query-server	6.0.0	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.276-amzn-0	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.276-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.276-amzn-0	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
trino-coordinator	398-amzn-0	Service zur Annahme von Abfragen und Verwaltung der Abfrageausführung der Trino-Worker.
trino-worker	398-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
trino-client	398-amzn-0	Trino-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Trino-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.

Komponente	Version	Beschreibung
r	4.0.2	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	2.0.0	Apache Ranger Key Management System
spark-client	3.3.0-amzn-1.1	Spark-Befehlszeilen-Clients.
spark-history-server	3.3.0-amzn-1.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	3.3.0-amzn-1.1	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	3.3.0-amzn-1.1	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
spark-rapids	22.08.0-amzn-0	Nvidia-Spark-RAPIDS-Plugin, das Apache Spark mit GPUs beschleunigt.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	2.10.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.10.2-amzn-0.1	Die Tez-YARN-Anwendung und -Bibliotheken.

Komponente	Version	Beschreibung
webserver	2.4.41+	Apache HTTP-Server.
zeppelin-server	0.10.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.5.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.5.10	ZooKeeper Befehlszeilen-Cli-ent.

## 6.9.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Aktionen zur Neukonfiguration treten auf, wenn Sie eine Konfiguration für Instance-Gruppen in einem laufenden Cluster angeben. Amazon EMR initiiert nur Rekonfigurationsaktionen für die Klassifizierungen, die Sie ändern. Weitere Informationen finden Sie unter [Eine Instance-Gruppe in einem laufenden Cluster neu konfigurieren](#).

### emr-6.9.1-Klassifizierungen

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.	Restarts the Resource Manager service.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
container-executor	Ändern Sie die Werte in der Datei „container-executor.cfg“ Datei von Hadoop YARN.	Not available.
container-log4j	Ändert die Werte in der container-log4j.properties-Datei in Hadoop YARN.	Not available.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Ändern Sie die Docker-bezogenen Einstellungen.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
emrfs-site	Ändert die EMRFS-Einstellungen.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegistrator, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.	Restarts Flink history server.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.	Restarts Flink history server.
flink-log4j-session	Ändern Sie die Einstellungen von Flink log4j-session.properties für Kubernetes-/Yarn-Sitzungen.	Restarts Flink history server.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.	Restarts Flink history server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.	Not available.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.	Not available.



Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Ändert die Werte in der HBase-Umgebung.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.	Not available.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.	This classification should not be reconfigured.
hdfs-env	Ändert die Werte in der HDFS-Umgebung.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei proto-hive-site .xml von HCatalog.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.	Restarts Hive WebHCat server.
hive	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Hive.	Sets configurations to launch Hive LLAP service.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.	Not available.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.	Not available.
hive-env	Ändert die Werte in der Hive-Umgebung.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.	Not available.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log	Not available.
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.	Not available.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hue-ini	Ändert die Werte in der INI-Datei in Hue.	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.	Restarts Hadoop Httpfs service.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.	Not available.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.	Not available.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Ändern der Werte in der Hudi-Umgebung.	Not available.
hudi-defaults	Ändern Sie die Werte in der hudi-defaults.conf-Datei in Hudi.	Not available.
iceberg-defaults	Ändern Sie die Werte in der iceberg-defaults.conf-Datei von Iceberg.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
delta-defaults	Ändern Sie die Werte in der delta-defaults.conf-Datei von Delta.	Not available.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.	Not available.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.	Not available.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.	Not available.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.	Not available.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.	Restarts Livy Server.
livy-env	Ändert die Werte in der Livy-Umgebung.	Restarts Livy Server.
livy-log4j2	Ändern Sie die log4j2.properties-Einstellungen für Livy.	Restarts Livy Server.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.	Restarts Hadoop MapReduce-HistoryServer.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
oozie-env	Ändert die Werte in der Oozie-Umgebung.	Restarts Oozie.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.	Restarts Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.	Not available.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.	Not available.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.	Not available.
pig-env	Ändert die Werte in der Pig-Umgebung.	Not available.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.	Restarts Oozie.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.	Not available.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-config	Ändert die Werte in der config.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.	Not available.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.	Not available.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.	Not available.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.	Not available.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.	Not available.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.	Not available.
presto-connector-lakeformation	Ändern Sie die Werte in der lakeformation.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.	Not available.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.	Not available.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.	Not available.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.	Not available.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.	Not available.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.	Not available.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.	Not available.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.	Not available.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.	Not available.



Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.	Not available.
trino-log	Ändern Sie die Werte in der log.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-config	Ändern Sie die Werte in der config.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Ändern Sie die Werte in der Trino-Datei password-authenticator.properties.	Restarts Trino-Server (for Trino)
trino-env	Ändern Sie die Werte in der trino-env.sh-Datei von Trino.	Restarts Trino-Server (for Trino)
trino-node	Ändern Sie die Werte in der node.properties-Datei in Trino.	Not available.
trino-connector-blackhole	Ändern Sie die Werte in der blackhole.properties-Datei in Trino.	Not available.
trino-connector-cassandra	Ändern Sie die Werte in der cassandra.properties-Datei in Trino.	Not available.
trino-connector-delta	Ändern Sie die Werte in der delta.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-hive	Ändern Sie die Werte in der hive.properties-Datei in Trino.	Restarts Trino-Server (for Trino)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-exchange-manager	Ändern Sie die Werte in der exchange-manager.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Ändern Sie die Werte in der iceberg.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Ändern Sie die Werte in der jmx.properties-Datei in Trino.	Not available.
trino-connector-kafka	Ändern Sie die Werte in der kafka.properties-Datei in Trino.	Not available.
trino-connector-localfile	Ändern Sie die Werte in der localfile.properties-Datei in Trino.	Not available.
trino-connector-memory	Ändern Sie die Werte in der memory.properties-Datei in Trino.	Not available.
trino-connector-mongodb	Ändern Sie die Werte in der mongodb.properties-Datei in Trino.	Not available.
trino-connector-mysql	Ändern Sie die Werte in der mysql.properties-Datei in Trino.	Not available.
trino-connector-postgresql	Ändern Sie die Werte in der postgresql.properties-Datei in Trino.	Not available.
trino-connector-raptor	Ändern Sie die Werte in der raptor.properties-Datei in Trino.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-connector-redis	Ändern Sie die Werte in der redis.properties-Datei in Trino.	Not available.
trino-connector-redshift	Ändern Sie die Werte in der redshift.properties-Datei in Trino.	Not available.
trino-connector-tpch	Ändern Sie die Werte in der tpch.properties-Datei in Trino.	Not available.
trino-connector-tpcds	Ändern Sie die Werte in der tpcds.properties-Datei in Trino.	Not available.
ranger-kms-dbks-site	Ändert die Werte in der dbks-site.xml-Datei von Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Ändert die Werte in der Ranger KMS-Umgebung.	Restarts Ranger KMS Server.
ranger-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei von Ranger KMS.	Not available.
ranger-kms-db-ca	Ändert die Werte für die CA-Datei auf S3 für die MySQL SSL-Verbindung mit Ranger KMS.	Not available.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.	This property modifies spark-defaults. See actions there.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Ändert die Werte in der Spark-Umgebung.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.	Not available.
spark-log4j2	Ändern Sie die Werte in der log4j2.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.	Not available.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.	Not available.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.	Not available.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.	Restart Oozie and HiveServer2.
yarn-env	Ändert die Werte in der YARN-Umgebung.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.	Restarts Zeppelin.
zeppelin-site	Ändern Sie die Konfigurationseinstellungen in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.	Restarts Zookeeper server.

## 6.9.1 Änderungsprotokoll

### Änderungsprotokoll für Version 6.9.1 und Versionshinweise

Datum	Ereignis	Beschreibung
2023-08-30	Versionshinweise der Aktualisierung	Den Versionshinweisen wurden mehrere Korrekturen im Zusammenhang mit der Steuerebene hinzugefügt
2023-08-21	Veröffentlichung von Dokumenten	Erste Veröffentlichung der Versionshinweise zu Amazon EMR 6.9.1

Datum	Ereignis	Beschreibung
2023-08-16	Die Bereitstellung ist abgeschlossen	<a href="#">Amazon EMR 6.9.1 vollständig in allen unterstützten Regionen bereitgestellt</a>
2023-08-04	Erstversion	Amazon EMR 6.9.1 wurde zuerst in begrenzten kommerziellen Regionen bereitgestellt

## Amazon-EMR-Version 6.9.0

### 6.9.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Delta](#), [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

### Informationen zur Anwendungsversion

	emr-6.9.0	emr-6.8.1	emr-6.8.0	emr-6.7.0
AWS SDK for Java	1.12.170	1.12,170	1.12,170	1.12,170

	emr-6.9.0	emr-6.8.1	emr-6.8.0	emr-6.7.0
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2.12,15	2.12,15	2.12,15	2.12,15
AmazonCloudWatchAgent	-	-	-	-
Delta	2.1.0	-	-	-
Flink	1.15.2	1.15.1	1.15.1	1.14.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4.13	2.4.12	2.4.12	2.4.4
HCatalog	3.1.3	3.1.3	3.1.3	3.1.3
Hadoop	3.3.3	3.2.1	3.2.1	3.2.1
Hive	3.1.3	3.1.3	3.1.3	3.1.3
Hudi	0.12.1-amzn-0	0.11.1-amzn-0	0.11.1-amzn-0	0.11.0-amzn-0
Hue	4.10.0	4.10.0	4.10.0	4.10.0
Iceberg	0.14.1-amzn-0	0.14.0-amzn-0	0.14.0-amzn-0	0.13.1-amzn-0
JupyterEnterpriseGateway	2.6.0	2.1.0	2.1.0	2.1.0
JupyterHub	1.4.1	1.4.1	1.4.1	1.4.1
Livy	0.7.1	0.7.1	0.7.1	0.7.1
MXNet	1.9.1	1.9.1	1.9.1	1.8.0
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1

	emr-6.9.0	emr-6.8.1	emr-6.8.0	emr-6.7.0
Phoenix	5.1.2	5.1.2	5.1.2	5.1.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0,276	0,273	0,273	0,272
Spark	3.3.0	3.3.0	3.3.0	3.2.1
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.10.0	2,9,1	2.9.1	2.4.1
Tez	0.10.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	398	388	388	378
Zeppelin	0,10,1	0.10.1	0.10.1	0.10.0
ZooKeeper	3.5.10	3.5.10	3.5.10	3.5.7

## 6.9.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 6.9.0. Änderungen beziehen sich auf Amazon-EMR-Version 6.8.0. Informationen zum Zeitplan der Veröffentlichung finden Sie im [Änderungsprotokoll](#).

### Neue Features

- Amazon EMR Version 6.9.0 unterstützt Apache Spark RAPIDS 22.08.0, Apache Hudi 0.12.1, Apache Iceberg 0.14.1, Trino 398 und Tez 0.10.2.
- Amazon EMR Version 6.9.0 enthält eine neue Open-Source-Anwendung, [Delta Lake](#) 2.1.0.
- Die Amazon-Redshift-Integration für Apache Spark ist in den Amazon-EMR-Versionen 6.9.0 und höher enthalten. Die native Integration war bisher ein Open-Source-Tool und ist ein Spark-Konnektor, mit dem Sie Apache-Spark-Anwendungen erstellen können, die Daten in Amazon Redshift und Amazon Redshift Serverless lesen und in diese schreiben. Weitere Informationen finden Sie unter [Amazon-Redshift-Integration für Apache Spark mit Amazon EMR verwenden](#).



- Amazon EMR Version 6.9.0 bietet Unterstützung für die Archivierung von Protokollen in Amazon S3 während der Cluster-Herunterskalierung. Bisher konnten Sie Protokolldateien nur während der Clusterbeendigung in Amazon S3 archivieren. Die neue Funktion stellt sicher, dass auf dem Cluster generierte Protokolldateien auch nach dem Beenden des Knotens auf Amazon S3 bestehen bleiben. Weitere Informationen finden Sie unter [Konfigurieren der Cluster-Protokollierung und des Debuggings](#).
- Um Abfragen mit langer Laufzeit zu unterstützen, verfügt Trino jetzt über einen fehlertoleranten Ausführungsmechanismus. Die fehlertolerante Ausführung minimiert Abfragefehler, indem fehlgeschlagene Abfragen oder deren Komponentenaufgaben wiederholt werden. Weitere Informationen finden Sie unter [Fehlertolerante Ausführung in Trino](#).
- Sie können Apache Flink auf Amazon EMR für die einheitliche BATCH- und STREAM-Verarbeitung von Apache-Hive-Tabellen oder Metadaten beliebiger Flink-Tabellenquellen wie Iceberg, Kinesis oder Kafka verwenden. Sie können den AWS Glue-Datenkatalog mithilfe der Amazon EMR-API, der AWS Management Console, der AWS CLI, oder als Metastore für Flink angeben. Weitere Informationen finden Sie unter [Konfiguration von Flink in Amazon EMR](#).
- Sie können jetzt AWS Identity and Access Management (IAM-) Runtime-Rollen und eine AWS Lake Formation basierte Zugriffskontrolle für Apache Spark-, Apache Hive- und Presto-Abfragen auf Amazon EMR auf EC2-Clustern mit Amazon Studio angeben. SageMaker Weitere Informationen finden Sie unter [Laufzeit-Rollen für Amazon EMR konfigurieren](#).

## Bekannte Probleme

- Für Amazon-EMR-Version 6.9.0 funktioniert Trino nicht auf Clustern, die für Apache Ranger aktiviert sind. Wenn Sie Trino mit Ranger verwenden müssen, wenden Sie sich an [AWS Support](#).
- Wenn Sie die Amazon-Redshift-Integration für Apache Spark verwenden und eine Zeit, `timetz`, `timestamp` oder `timestampz` mit Mikrosekundengenauigkeit im Parquet-Format haben, rundet der Konnektor die Zeitwerte auf den nächstliegenden Millisekundenwert. Um das Problem zu umgehen, verwenden Sie den `unload_s3_format`-Formatparameter-Text-Unload.
- Wenn Sie Spark mit der Formatierung des Hive-Partitionsstandorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den Amazon-EMR-Versionen 5.30.0 bis 5.36.0 und 6.2.0 bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:
  - Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.

- Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist `s3://bucket/table/p=a` ein Präfix von `s3://bucket/table/p=a b`.
- Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen UTF-8-Wert, der kleiner als das `/`-Zeichen (U+002F) ist. Beispielsweise fällt das Leerzeichen (U+0020), das in `s3://bucket/table/p=a b` zwischen `a` und `b` vorkommt, in diese Kategorie. Beachten Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: `!"#$%&'()*+,-.` Weitere Informationen finden Sie unter [UTF-8-Kodierungstabelle und Unicode-Zeichen](#).

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.

- Verbindungen zu Amazon EMR-Clustern von Amazon SageMaker Studio aus können zeitweise mit dem Antwortcode 403 Forbidden fehlschlagen. Dieser Fehler tritt auf, wenn die Einrichtung der IAM-Rolle auf dem Cluster länger als 60 Sekunden dauert. Um dieses Problem zu umgehen, können Sie einen Amazon-EMR-Patch installieren, um Wiederholungsversuche zu ermöglichen und das Timeout auf mindestens 300 Sekunden zu erhöhen. Verwenden Sie die folgenden Schritte, um die Bootstrap-Aktion anzuwenden, wenn Sie Ihren Cluster starten.

1. Laden Sie das Bootstrap-Skript und die RPM-Dateien von den folgenden Amazon-S3-URIs herunter.

```
s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/gcsc/replace-rpms.sh
s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/gcsc/emr-secret-agent-1.18.0-SNAPSHOT20221121212949.noarch.rpm
```

2. Laden Sie die Dateien aus dem vorherigen Schritt in einen Amazon-S3-Bucket hoch, dessen Eigentümer Sie sind. Der Bucket muss sich in dem Bereich befinden, in AWS-Region dem Sie den Cluster starten möchten.
3. Schließen Sie die folgende Bootstrap-Aktion ein, wenn Sie Ihren EMR-Cluster starten. Ersetzen Sie `bootstrap_URI` und `RPM_URI` durch die entsprechenden URIs von Amazon S3.

```
--bootstrap-actions "Path=bootstrap_URI,Args=[RPM_URI]"
```

- Bei den Amazon-EMR-Versionen 5.36.0 und 6.6.0 bis 6.9.0 kann es bei `SecretAgent`- und `RecordServer`-Servicekomponenten aufgrund einer falschen Konfiguration des

Dateinamenmusters in den Log4j2-Eigenschaften zu einem Verlust von Protokolldaten kommen. Die falsche Konfiguration führt dazu, dass die Komponenten nur eine Protokolldatei pro Tag generieren. Wenn die Rotationsstrategie angewendet wird, überschreibt sie die vorhandene Datei, anstatt wie erwartet eine neue Protokolldatei zu generieren. Um das Problem zu umgehen, verwenden Sie eine Bootstrap-Aktion, um jede Stunde Protokolldateien zu generieren, und fügen Sie eine automatisch inkrementierte Ganzzahl an den Dateinamen an, um die Rotation zu handhaben.

Verwenden Sie für die Versionen Amazon EMR 6.6.0 bis 6.9.0 die folgende Bootstrap-Aktion, wenn Sie einen Cluster starten.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-6x/replace-puppet.sh,Args=[]"
```

Verwenden Sie für Amazon EMR 5.36.0 die folgende Bootstrap-Aktion, wenn Sie einen Cluster starten.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-5x/replace-puppet.sh,Args=[]"
```

- Apache Flink bietet native S3 FileSystem - und FileSystem Hadoop-Konnektoren, mit denen Anwendungen Daten erstellen FileSink und in Amazon S3 schreiben können. Dies FileSink schlägt mit einer der folgenden zwei Ausnahmen fehl.

```
java.lang.UnsupportedOperationException: Recoverable writers on Hadoop are only supported for HDFS
```

```
Caused by: java.lang.NoSuchMethodError:  
  org.apache.hadoop.io.retry.RetryPolicies.retryOtherThanRemoteAndSaslException(Lorg/  
apache/hadoop/io/retry/RetryPolicy;Ljava/util/Map;)Lorg/apache/hadoop/io/retry/  
RetryPolicy;  
                                     at  
  org.apache.hadoop.yarn.client.RMProxy.createRetryPolicy(RMProxy.java:302) ~[hadoop-  
yarn-common-3.3.3-amzn-0.jar:?]
```

Als Workaround können Sie einen Amazon-EMR-Patch installieren, der das oben genannte Problem in Flink behebt. Führen Sie die folgenden Schritte aus, um die Bootstrap-Aktion beim Start Ihres Clusters anzuwenden.

1. Laden Sie das flink-rpm in Ihren Amazon-S3-Bucket herunter. Ihr RPM-Pfad ist `s3://DOC-EXAMPLE-BUCKET/rpms/flink/`.
2. Laden Sie das Bootstrap-Skript und die RPM-Dateien mit dem folgenden URI von Amazon S3 herunter. `regionName` Ersetzen Sie es durch den AWS-Region Ort, an dem Sie den Cluster starten möchten.

```
s3://emr-data-access-control-regionName/customer-bootstrap-actions/gcsc/replace-rpms.sh
```

3. Mit Hadoop 3.3.3 wurde eine Änderung in YARN ([YARN-9608](#)) eingeführt, die dafür sorgt, dass Knoten, auf denen Container laufen, außer Betrieb genommen werden, bis die Anwendung abgeschlossen ist. Diese Änderung stellt sicher, dass lokale Daten wie Shuffle-Daten nicht verloren gehen und Sie den Auftrag nicht erneut ausführen müssen. In Amazon EMR 6.8.0 und 6.9.0 kann dieser Ansatz auch zu einer Unterauslastung von Ressourcen auf Clustern mit oder ohne aktivierter verwalteter Skalierung führen.

Mit [Amazon EMR 6.10.0](#) gibt es einen Workaround für dieses Problem, indem Sie den Wert in `yarn-site.xml` von `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` auf `false` setzen. In den Amazon-EMR-Versionen 6.11.0 und höher sowie 6.8.1, 6.9.1 und 6.10.1 ist die Konfiguration standardmäßig auf `false` eingestellt, um dieses Problem zu beheben.

## Änderungen, Verbesserungen und behobene Probleme

- Für Amazon-EMR-Version 6.9.0 und höher verwenden alle von Amazon EMR installierten Komponenten, die Log4j-Bibliotheken verwenden, Log4j Version 2.17.1 oder höher.
- Wenn Sie den DynamoDB-Konnektor mit Spark auf den Amazon-EMR-Versionen 6.6.0, 6.7.0 und 6.8.0 verwenden, geben alle Lesevorgänge aus Ihrer Tabelle ein leeres Ergebnis zurück, obwohl der Eingabe-Split auf nicht leere Daten verweist. Amazon EMR Version 6.9.0 behebt dieses Problem.
- Amazon EMR 6.9.0 bietet eingeschränkte Unterstützung für die auf Lake Formation basierende Zugriffskontrolle mit Apache Hudi beim Lesen von Daten mit Spark SQL. Die Unterstützung gilt für SELECT-Abfragen mit Spark SQL und ist auf die Zugriffskontrolle auf Spaltenebene beschränkt. Weitere Informationen finden Sie unter [Hudi und Lake Formation](#).

- Wenn Sie Amazon EMR 6.9.0 verwenden, um einen Hadoop-Cluster mit aktivierten [Knoten-Kennzeichnungen](#) zu erstellen, gibt die [YARN-Metriken-API](#) aggregierte Informationen über alle Partitionen statt der Standardpartition zurück. Weitere Informationen finden Sie unter [YARN-11414](#).
- Mit Amazon-EMR-Version 6.9.0 haben wir Trino auf Version 398 aktualisiert, die Java 17 verwendet. Die vorherige unterstützte Version von Trino für Amazon EMR 6.8.0 war Trino 388, das auf Java 11 lief. Weitere Informationen zu dieser Änderung finden Sie unter [Trino-Updates für Java 17 im Trino-Blog](#).
- Diese Version behebt ein Problem mit der Nichtübereinstimmung der Timing-Sequenz zwischen Apache BigTop und Amazon EMR in der EC2-Cluster-Startsequenz. Diese Nichtübereinstimmung der Timing-Sequenz tritt auf, wenn ein System versucht, zwei oder mehr Operationen gleichzeitig auszuführen, anstatt sie in der richtigen Reihenfolge auszuführen. Infolgedessen kam es bei bestimmten Clusterkonfigurationen zu Timeouts beim Instance-Start und zu langsameren Cluster-Startup-Zeiten.
- Wenn Sie einen Cluster mit der neuesten Patch-Version von Amazon EMR 5.36 oder höher, 6.6 oder höher oder 7.0 oder höher starten, verwendet Amazon EMR die neueste Version von Amazon Linux 2023 oder Amazon Linux 2 für das standardmäßige Amazon EMR-AMI. Weitere Informationen finden Sie unter [Verwenden des standardmäßigen Amazon-Linux-AMI für Amazon EMR](#).

#### Note

Diese Version erhält keine automatischen AMI-Updates mehr, da sie von einer weiteren Patch-Versionen abgelöst wurde. Die Patch-Version wird durch die Zahl nach dem zweiten Dezimalpunkt (6.8.**1**) gekennzeichnet. Um zu sehen, ob Sie die neueste Patch-Version verwenden, überprüfen Sie die verfügbaren Versionen im [Versionshandbuch](#) oder überprüfen Sie die Amazon-EMR-Versions-Dropdownliste, wenn Sie einen Cluster in der Konsole erstellen, oder verwenden Sie die [ListReleaseLabels](#)-API- oder [list-release-labels](#)-CLI-Aktion. Um aktuelle Informationen über Neuerscheinungen erhalten, abonnieren Sie den RSS-Feed auf der Seite [Was ist neu?](#).

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.20230808.0	4,14.320	24. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 727.0	4,14.320	14. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 719,0	4,14.320	02. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv)



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 628.0	4,14.318	12. Juli 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 612,0	4,14.314	23. Juni 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 504.1	4,14.313	16. Mai 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 418,0	4,14.311	3. Mai 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 404.1	4,14.311	18. April 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)
2.0.2023 404.0	4,14.311	10. April 2023	USA Ost (Nord-Virginia), Europa (Paris)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 320.0	4,14.309	30. März 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 307,0	4,14.305	15. März 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202307,0	4,14.304	22. Februar 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022 210.1	4.14.301	12. Januar 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022103.3	4,14.296	5. Dezember 2022	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

## 6.9.0 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version

sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-amzn-*EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen myapp-component der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als 2.2-amzn-2 aufgeführt.

Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1.4.2	Amazon SageMaker Spark-SDK
delta	2.1.0	Delta Lake ist ein offenes Tabellenformat für riesige analytische Datensätze
emr-ddb	4.16.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	3.3.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.6.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-notebook-env	1.7.0	Conda Env für EMR-Notebooks, das Jupyter Enterprise Gateway enthält
emr-s3-dist-cp	2.23.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	2.2.0	EMR S3Select-Konnektor

Komponente	Version	Beschreibung
emrfs	2,54,0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.15.2	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
flink-jobmanager-config	1.15.2	Verwaltung von Ressourcen auf EMR-Knoten für Apache JobManager Flink.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	3.3.3-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	3.3.3-amzn-1	HDFS-Service auf Knotenebene zum Speichern von Blöcken.

Komponente	Version	Beschreibung
hadoop-hdfs-library	3.3.3-amzn-1	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	3.3.3-amzn-1	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-hdfs-journalnode	3.3.3-amzn-1	HDFS-Service zum Verwalten des Hadoop-Dateisystem journals auf HA-Clustern.
hadoop-httpfs-server	3.3.3-amzn-1	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	3.3.3-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	3.3.3-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	3.3.3-amzn-1	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resource-manager	3.3.3-amzn-1	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.

Komponente	Version	Beschreibung
hadoop-yarn-timeline-server	3.3.3-amzn-1	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	2.4.13-amzn-0	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	2.4.13-amzn-0	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	2.4.13-amzn-0	HBase-Befehlszeilen-Client.
hbase-rest-server	2.4.13-amzn-0	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	2.4.13-amzn-0	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hbase-operator-tools	2.4.13-amzn-0	Reparaturtool für Apache-HBase-Cluster.
hcatalog-client	3.1.3-amzn-2	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	3.1.3-amzn-2	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).

Komponente	Version	Beschreibung
hcatalog-webhcat-server	3.1.3-amzn-2	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	3.1.3-amzn-2	Hive-Befehlszeilen-Client.
hive-hbase	3.1.3-amzn-2	Hive-hbase client.
hive-metastore-server	3.1.3-amzn-2	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	3.1.3-amzn-2	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.12.1-amzn-0	Inkrementelles Verarbeitungs-Framework zur Versorgung der Datenpipeline mit geringer Latenz und hoher Effizienz.
hudi-presto	0.12.1-amzn-0	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hudi-trino	0.12.1-amzn-0	Bündel-Bibliothek zum Ausführen von Trino mit Hudi.
hudi-spark	0.12.1-amzn-0	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.
hue-server	4.10.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.

Komponente	Version	Beschreibung
iceberg	0.14.1-amzn-0	Apache Iceberg ist ein offenes Tabellenformat für sehr große analytische Datensätze
jupyterhub	1.4.1	Multi-User-Server für Jupyter-Notebooks
livy-server	0.7.1-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP- und Reverse-Proxy-Server.
mxnet	1.9.1	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.68+	MariaDB-Datenbankserver.
nvidia-cuda	11,7,0	Nvidia-Treiber und Cuda-Toolkit
oozie-client	5.2.1	Oozie-Befehlszeilen-Client.
oozie-server	5.2.1	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	5.1.2	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-connectors	6.0.0-SNAPSHOT	Apache-Phoenix-Konnektoren für Spark-3



Komponente	Version	Beschreibung
phoenix-query-server	6.0.0	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.276-amzn-0	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.276-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.276-amzn-0	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
trino-coordinator	398-amzn-0	Service zur Annahme von Abfragen und Verwaltung der Abfrageausführung der Trino-Worker.
trino-worker	398-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
trino-client	398-amzn-0	Trino-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Trino-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.

Komponente	Version	Beschreibung
r	4.0.2	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	2.0.0	Apache Ranger Key Management System
spark-client	3.3.0-amzn-1	Spark-Befehlszeilen-Clients.
spark-history-server	3.3.0-amzn-1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	3.3.0-amzn-1	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	3.3.0-amzn-1	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
spark-rapids	22.08.0-amzn-0	Nvidia-Spark-RAPIDS-Plugin, das Apache Spark mit GPUs beschleunigt.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	2.10.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.10.2-amzn-0	Die Tez-YARN-Anwendung und -Bibliotheken.

Komponente	Version	Beschreibung
webserver	2.4.41+	Apache HTTP-Server.
zeppelin-server	0.10.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.5.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.5.10	ZooKeeper Befehlszeilen-Cli-ent.

## 6.9.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Aktionen zur Neukonfiguration treten auf, wenn Sie eine Konfiguration für Instance-Gruppen in einem laufenden Cluster angeben. Amazon EMR initiiert nur Rekonfigurationsaktionen für die Klassifizierungen, die Sie ändern. Weitere Informationen finden Sie unter [Eine Instance-Gruppe in einem laufenden Cluster neu konfigurieren](#).

### emr-6.9.0-Klassifizierungen

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.	Restarts the Resource Manager service.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
container-executor	Ändern Sie die Werte in der Datei „container-executor.cfg“ Datei von Hadoop YARN.	Not available.
container-log4j	Ändert die Werte in der container-log4j.properties-Datei in Hadoop YARN.	Not available.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Ändern Sie die Docker-bezogenen Einstellungen.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
emrfs-site	Ändert die EMRFS-Einstellungen.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegistrator, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.	Restarts Flink history server.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.	Restarts Flink history server.
flink-log4j-session	Ändern Sie die Einstellungen von Flink log4j-session.properties für Kubernetes-/Yarn-Sitzungen.	Restarts Flink history server.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.	Restarts Flink history server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.	Not available.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Ändert die Werte in der HBase-Umgebung.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.	Not available.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.	This classification should not be reconfigured.
hdfs-env	Ändert die Werte in der HDFS-Umgebung.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei proto-hive-site .xml von HCatalog.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.	Restarts Hive WebHCat server.
hive	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Hive.	Sets configurations to launch Hive LLAP service.



Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.	Not available.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.	Not available.
hive-env	Ändert die Werte in der Hive-Umgebung.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.	Not available.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log	Not available.
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.	Not available.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hue-ini	Ändert die Werte in der INI-Datei in Hue.	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.	Restarts Hadoop Httpfs service.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.	Not available.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.	Not available.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Ändern der Werte in der Hudi-Umgebung.	Not available.
hudi-defaults	Ändern Sie die Werte in der hudi-defaults.conf-Datei in Hudi.	Not available.
iceberg-defaults	Ändern Sie die Werte in der iceberg-defaults.conf-Datei von Iceberg.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
delta-defaults	Ändern Sie die Werte in der delta-defaults.conf-Datei von Delta.	Not available.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.	Not available.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.	Not available.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.	Not available.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.	Not available.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.	Restarts Livy Server.
livy-env	Ändert die Werte in der Livy-Umgebung.	Restarts Livy Server.
livy-log4j2	Ändern Sie die log4j2.properties-Einstellungen für Livy.	Restarts Livy Server.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.	Restarts Hadoop MapReduce-HistoryServer.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
oozie-env	Ändert die Werte in der Oozie-Umgebung.	Restarts Oozie.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.	Restarts Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.	Not available.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.	Not available.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.	Not available.
pig-env	Ändert die Werte in der Pig-Umgebung.	Not available.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.	Restarts Oozie.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.	Not available.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-config	Ändert die Werte in der config.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.	Not available.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.	Not available.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.	Not available.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.	Not available.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.	Not available.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.	Not available.
presto-connector-lakeformation	Ändern Sie die Werte in der lakeformation.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.	Not available.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.	Not available.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.	Not available.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.	Not available.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.	Not available.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.	Not available.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.	Not available.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.	Not available.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.	Not available.
trino-log	Ändern Sie die Werte in der log.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-config	Ändern Sie die Werte in der config.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Ändern Sie die Werte in der Trino-Datei password-authenticator.properties.	Restarts Trino-Server (for Trino)
trino-env	Ändern Sie die Werte in der trino-env.sh-Datei von Trino.	Restarts Trino-Server (for Trino)
trino-node	Ändern Sie die Werte in der node.properties-Datei in Trino.	Not available.
trino-connector-blackhole	Ändern Sie die Werte in der blackhole.properties-Datei in Trino.	Not available.
trino-connector-cassandra	Ändern Sie die Werte in der cassandra.properties-Datei in Trino.	Not available.
trino-connector-delta	Ändern Sie die Werte in der delta.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-hive	Ändern Sie die Werte in der hive.properties-Datei in Trino.	Restarts Trino-Server (for Trino)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-exchange-manager	Ändern Sie die Werte in der exchange-manager.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Ändern Sie die Werte in der iceberg.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Ändern Sie die Werte in der jmx.properties-Datei in Trino.	Not available.
trino-connector-kafka	Ändern Sie die Werte in der kafka.properties-Datei in Trino.	Not available.
trino-connector-localfile	Ändern Sie die Werte in der localfile.properties-Datei in Trino.	Not available.
trino-connector-memory	Ändern Sie die Werte in der memory.properties-Datei in Trino.	Not available.
trino-connector-mongodb	Ändern Sie die Werte in der mongodb.properties-Datei in Trino.	Not available.
trino-connector-mysql	Ändern Sie die Werte in der mysql.properties-Datei in Trino.	Not available.
trino-connector-postgresql	Ändern Sie die Werte in der postgresql.properties-Datei in Trino.	Not available.
trino-connector-raptor	Ändern Sie die Werte in der raptor.properties-Datei in Trino.	Not available.



Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-connector-redis	Ändern Sie die Werte in der redis.properties-Datei in Trino.	Not available.
trino-connector-redshift	Ändern Sie die Werte in der redshift.properties-Datei in Trino.	Not available.
trino-connector-tpch	Ändern Sie die Werte in der tpch.properties-Datei in Trino.	Not available.
trino-connector-tpcds	Ändern Sie die Werte in der tpcds.properties-Datei in Trino.	Not available.
ranger-kms-dbks-site	Ändert die Werte in der dbks-site.xml-Datei von Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Ändert die Werte in der Ranger KMS-Umgebung.	Restarts Ranger KMS Server.
ranger-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei von Ranger KMS.	Not available.
ranger-kms-db-ca	Ändert die Werte für die CA-Datei auf S3 für die MySQL SSL-Verbindung mit Ranger KMS.	Not available.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.	This property modifies spark-defaults. See actions there.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Ändert die Werte in der Spark-Umgebung.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.	Not available.
spark-log4j2	Ändern Sie die Werte in der log4j2.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.	Not available.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.	Not available.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.	Not available.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.	Restart Oozie and HiveServer2.
yarn-env	Ändert die Werte in der YARN-Umgebung.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.	Restarts Zeppelin.
zeppelin-site	Ändern Sie die Konfigurationseinstellungen in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.	Restarts Zookeeper server.

## 6.9.0 Änderungsprotokoll

### Änderungsprotokoll für Version 6.9.0 und Versionshinweise

Datum	Ereignis	Beschreibung
2023-08-30	Versionshinweise der Aktualisierung	Korrektur für das Problem mit der Nichtübereinstimmung der Timing-Sequenz hinzugefügt
2023-08-21	Versionshinweise der Aktualisierung	Es wurde ein bekanntes Problem mit Hadoop 3.3.3 hinzugefügt.

Datum	Ereignis	Beschreibung
2023-07-26	Aktualisierung	Neue Betriebssystem-Release-Labels 2.0.20230612.0 und 2.0.20230628.0 .
13.12.2022	Versionshinweise	Funktion hinzugefügt und bekanntes Problem bei Runtime mit SageMaker
29.11.2022	Versionshinweise und Dokumentation wurden aktualisiert	Feature für die Amazon-Redshift-Integration für Apache Spark hinzugefügt
23.11.2022	Versionshinweise	Der Log4j-Eintrag wurde entfernt
18.11.2022	Die Bereitstellung ist abgeschlossen	<a href="#">Amazon EMR 6.9 vollständig in allen unterstützten Regionen bereitgestellt</a>
18.11.2022	Veröffentlichung von Dokumenten	Erste Veröffentlichung der Versionshinweise zu Amazon EMR 6.9
14.11.2022	Erstversion	Amazon EMR 6.9 wurde zuerst in begrenzten kommerziellen Regionen bereitgestellt

## Amazon-EMR-Version 6.8.1

### 6.8.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

#### Informationen zur Anwendungsversion

	emr-6.8.1	emr-6.8.0	emr-6.7.0	emr-6.6.0
AWS SDK for Java	1.12.170	1.12,170	1.12,170	1.12,170
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2.12,15	2.12,15	2.12,15	2.12,10
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.15.1	1.15.1	1.14.2	1.14.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4.12	2.4.12	2.4.4	2.4.4
HCatalog	3.1.3	3.1.3	3.1.3	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.3	3.1.3	3.1.3	3.1.2

	emr-6.8.1	emr-6.8.0	emr-6.7.0	emr-6.6.0
Hudi	0.11.1-amzn-0	0.11.1-amzn-0	0.11.0-amzn-0	0.10.1-amzn-0
Hue	4.10.0	4.10.0	4.10.0	4.10.0
Iceberg	0.14.0-amzn-0	0.14.0-amzn-0	0.13.1-amzn-0	0.13.1
JupyterEnterpriseGateway	2.1.0	2.1.0	2.1.0	2.1.0
JupyterHub	1.4.1	1.4.1	1.4.1	1.4.1
Livy	0.7.1	0.7.1	0.7.1	0.7.1
MXNet	1.9.1	1.9.1	1.8.0	1.8.0
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	5.1.2	5.1.2	5.1.2	5.1.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0,273	0,273	0,272	0,267
Spark	3.3.0	3.3.0	3.2.1	3.2.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2,9,1	2.9.1	2.4.1	2.4.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	388	388	378	367
Zeppelin	0,10,1	0.10.1	0.10.0	0.10.0
ZooKeeper	3.5.10	3.5.10	3.5.7	3.5.7

## 6.8.1 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 6.8.1. Änderungen beziehen sich auf Version 6.8.0. Informationen zum Zeitplan für die Veröffentlichung finden Sie unter [6.8.1 Änderungsprotokoll](#).

### Änderungen, Verbesserungen und behobene Probleme

- Mit Hadoop 3.3.3 wurde eine Änderung in YARN ([YARN-9608](#)) eingeführt, die dafür sorgt, dass Knoten, auf denen Container laufen, außer Betrieb genommen werden, bis die Anwendung abgeschlossen ist. Diese Änderung stellt sicher, dass lokale Daten wie Shuffle-Daten nicht verloren gehen und Sie den Auftrag nicht erneut ausführen müssen. Dieser Ansatz kann auch zu einer Unterauslastung von Ressourcen in Clustern mit oder ohne aktivierter verwalteter Skalierung führen.

Bei den Amazon-EMR-Versionen 6.11.0 und höher sowie 6.8.1, 6.9.1 und 6.10.1 ist der Wert in `yarn-site.xml` von `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` auf `false` gesetzt, um dieses Problem zu beheben.

Der Fix behebt zwar die Probleme, die durch YARN-9608 eingeführt wurden, kann jedoch dazu führen, dass Hive-Aufträge aufgrund von Shuffle-Datenverlusten auf Clustern, für die verwaltete Skalierung aktiviert ist, fehlschlagen. Wir haben dieses Risiko in dieser Version verringert, indem wir auch `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-shuffle-data` für Hive-Workloads eingerichtet haben. Diese Konfiguration ist nur mit Amazon-EMR-Versionen 6.11.0 und höher verfügbar.

- Der Metrics Collector sendet nach einem Failover des Primärknotens in Clustern mit der Instance-Gruppenkonfiguration keine Metriken an die Steuerebene.
- Diese Version macht Wiederholungsversuche bei fehlgeschlagenen HTTP-Anfragen an die Endpunkte der Metrics Collector überflüssig.
- Diese Version enthält eine Änderung, die es Hochverfügbarkeitsclustern ermöglicht, nach einem Neustart den Status „Fehler“ wiederherzustellen.
- Diese Version behebt ein Problem, bei dem große, von Benutzern erstellte UUIDs zu Überlaufausnahmen führten.
- Diese Version behebt Timeout-Probleme beim Amazon-EMR-Rekonfigurationsprozess.
- Diese Version verhindert ein Problem, bei dem eine fehlgeschlagene Neukonfiguration andere, nicht miteinander verbundene Prozesse unterbrechen könnte.
- Diese Version enthält Sicherheitskorrekturen.

- Diese Version behebt ein Problem, bei dem Cluster, die Workloads auf Spark mit Amazon EMR ausführen, möglicherweise unbemerkt falsche Ergebnisse mit `contains`, `startsWith`, `endsWith` und `like` erhalten. Dieses Problem tritt auf, wenn Sie die Ausdrücke für partitionierte Felder verwenden, die Metadaten im Amazon EMR Hive3 Metastore Server (HMS) haben.
- Bei Amazon EMR 6.6.0 bis 6.9.x haben INSERT-Abfragen mit dynamischer Partition und einer ORDER-BY- oder SORT-BY-Klausel immer zwei Reduzierungen. Dieses Problem wird durch die OSS-Änderung [HIVE-20703](#) verursacht, die die Optimierung dynamischer Sortierpartitionen einer kostenbasierten Entscheidung unterstellt. Wenn Ihr Workload keine Sortierung dynamischer Partitionen erfordert, empfehlen wir Ihnen, die `hive.optimize.sort.dynamic.partition.threshold`-Eigenschaft auf -1 zu setzen, um das neue Feature zu deaktivieren und die korrekt berechnete Anzahl von Reduzierern zu erhalten. Dieses Problem wurde behoben in OSS Hive als Teil von [HIVE-22269](#) und in Amazon EMR 6.10.0 behoben.
- In Hive kann es zu Datenverlusten kommen, wenn Sie HDFS als Scratch-Verzeichnis verwenden und die Option „Kleine Dateien zusammenführen“ aktiviert haben und die Tabelle statische Partitionspfade enthält.
- Diese Version behebt ein Leistungsproblem mit Hive, wenn das Zusammenführen kleiner Dateien (standardmäßig deaktiviert) am Ende des ETL-Auftrags aktiviert ist.
- Diese Version behebt ein Problem mit der Drosselung auf der Glue-Seite, wenn es keine benutzerdefinierten Funktionen (UDF) gibt.
- Diese Version behebt ein Problem, bei dem Container-Protokolle vom Knotenprotokoll-Aggregationsservice gelöscht werden, bevor Log Pusher sie im Falle einer Außerbetriebnahme von YARN an S3 weiterleiten kann.
- Diese Version behebt den Umgang mit komprimierten/archivierten Dateien mit persistenter Speicherdateiverfolgung für HBase.
- Diese Version behebt ein Problem, das die Leistung von Spark beeinträchtigte, wenn Sie einen `true`-Standardwert für die `spark.yarn.heterogeneousExecutors.enabled`-Konfiguration in `spark-defaults.conf` festlegen.
- Diese Version behebt ein Problem, bei dem Reduce Task die Shuffle-Daten nicht lesen konnte. Das Problem führte zu Hive-Abfragefehlern mit einem Speicherfehler.
- Diese Version behebt ein Problem, das dazu führte, dass der Node Provisioner ausfiel, wenn der HDFS NameNode (NN) -Dienst während des Knotenaustauschs im Safemode hängen blieb.
- Diese Version fügt dem Cluster-Skalierungs-Workflow für EMR-Cluster, auf denen Presto oder Trino ausgeführt werden, einen neuen Wiederholungsmechanismus hinzu. Diese Verbesserung



verringert das Risiko, dass die Clustergrößenänderung aufgrund eines einzigen fehlgeschlagenen Größenänderungsvorgangs auf unbestimmte Zeit zum Stillstand kommt. Sie verbessert auch die Clusterauslastung, da Ihr Cluster schneller hoch- und herunterskaliert wird.

- Diese Version verbessert die Cluster-Herunterskalierungs-Logik, sodass Ihr Cluster nicht versucht, die Core-Knoten unter die für den Cluster eingestellte HDFS-Replikationsfaktor-Einstellung herunterzuskalieren. Dies entspricht Ihren Anforderungen an die Datenredundanz und verringert die Wahrscheinlichkeit, dass ein Skalierungsvorgang zum Stillstand kommt.
- Der Protokoll-Management-Daemon wurde aktualisiert, um alle Protokolle, die aktiv verwendet werden, mit offenen Datei-Handles auf dem lokalen Instance-Speicher und die zugehörigen Prozesse zu identifizieren. Dieses Upgrade stellt sicher, dass Amazon EMR die Dateien ordnungsgemäß löscht und Speicherplatz zurückgewinnt, nachdem die Protokolle in Amazon S3 archiviert wurden.
- Diese Version beinhaltet eine Erweiterung des Protokollverwaltungs-Daemons, die leere, unbenutzte Schrittverzeichnisse im lokalen Cluster-Dateisystem löscht. Eine zu große Anzahl leerer Verzeichnisse kann die Leistung der Amazon-EMR-Daemons beeinträchtigen und zu einer Überauslastung der Festplatte führen.
- Diese Version behebt ein Problem, das auftreten kann, wenn Sie einen Edge-Knoten erstellen, indem Sie einen der Primärknoten aus einem Cluster mit mehreren Primärknoten replizieren. Der replizierte Edge-Knoten kann zu Verzögerungen bei Scale-Down-Vorgängen oder zu einer hohen Arbeitsspeicherauslastung auf den Primärknoten führen. Weitere Informationen zum Erstellen eines Edge-Knotens für die Kommunikation mit Ihrem EMR-Cluster finden Sie unter [Edge Node Creator](#) im `aws-samples` Repo unter `GitHub`
- Diese Version verbessert den Automatisierungsprozess, den Amazon EMR verwendet, um Amazon-EBS-Volumes nach einem Neustart erneut in eine Instance einzubinden.
- Diese Version behebt ein Problem, das zu zeitweiligen Lücken in den Hadoop-Metriken führte, die Amazon EMR auf Amazon veröffentlicht. CloudWatch
- Diese Version behebt ein Problem mit EMR-Clustern, bei dem ein Update der YARN-Konfigurationsdatei, die die Ausschlussliste der Knoten für den Cluster enthält, aufgrund einer Überauslastung der Festplatte unterbrochen wird. Das unvollständige Update behindert zukünftige Cluster-Herunterskalierungs-Vorgänge. Diese Version stellt sicher, dass Ihr Cluster fehlerfrei bleibt und dass die Skalierungsvorgänge wie erwartet funktionieren.
- Diese Version verbessert den On-Cluster-Log-Management-Daemon zur Überwachung zusätzlicher Protokollordner in Ihrem EMR-Cluster. Durch diese Verbesserung werden Szenarien mit übermäßiger Festplattenauslastung minimiert.

- Diese Version startet den Protokoll-Management-Daemon auf dem Cluster automatisch neu, wenn er beendet wird. Durch diese Verbesserung wird das Risiko verringert, dass Knoten aufgrund übermäßiger Festplattenauslastung als fehlerhaft erscheinen.
- Diese Version bietet Unterstützung für die Archivierung von Protokollen in Amazon S3 während der Cluster-Scale-Down. Bisher konnten Sie Protokolldateien nur während der Clusterbeendigung in Amazon S3 archivieren. Die neue Funktion stellt sicher, dass auf dem Cluster generierte Protokolldateien auch nach dem Beenden des Knotens auf Amazon S3 bestehen bleiben. Weitere Informationen finden Sie unter [Konfigurieren der Cluster-Protokollierung und des Debuggings](#).
- Diese Version behebt ein Problem, das auftrat, als der Amazon S3 URI für eine Bootstrap-Aktion mit einer Portnummer endete, zum Beispiel: a . b . c . d : 4345. Amazon EMR analysierte diese URIs falsch, sodass alle zugehörigen Bootstrap-Aktionen fehlschlagen.
- Diese Version behebt ein Problem mit der Nichtübereinstimmung der Timing-Sequenz zwischen Apache BigTop und Amazon EMR in der EC2-Cluster-Startsequenz. Diese Nichtübereinstimmung der Timing-Sequenz tritt auf, wenn ein System versucht, zwei oder mehr Operationen gleichzeitig auszuführen, anstatt sie in der richtigen Reihenfolge auszuführen. Infolgedessen kam es bei bestimmten Clusterkonfigurationen zu Timeouts beim Instance-Start und zu langsameren Cluster-Startup-Zeiten.
- Wenn Sie einen Cluster mit der neuesten Patch-Version von Amazon EMR 5.36 oder höher, 6.6 oder höher oder 7.0 oder höher starten, verwendet Amazon EMR die neueste Version von Amazon Linux 2023 oder Amazon Linux 2 für das standardmäßige Amazon EMR-AMI. Weitere Informationen finden Sie unter [Verwenden des standardmäßigen Amazon-Linux-AMI für Amazon EMR](#).

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 223.0	4,14.336	8. März 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024131.0	4,14.336	14. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 124.0	4,14.336	7. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024109.0	4,14.334	24. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 218.0	4,14.330	2. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202306,0	4,14.330	22. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Kanada (Zentral), Israel AWS GovCloud (Tel Aviv), AWS GovCloud (US-West),



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023116.0	4,14.328	11. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Kanada (Zentral), Israel AWS GovCloud (Tel Aviv), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023101.0	4,14.327	16. November 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Kanada (Zentral), Israel AWS GovCloud (Tel Aviv), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023020.1	4,14.326	7. November 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Kanada (Zentral), Israel AWS GovCloud (Tel Aviv), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023012.1	4,14.326	26. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Kanada (Zentral), Israel AWS GovCloud (Tel Aviv), AWS GovCloud (US-West),



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 926.0	4,14.322	19. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Kanada (Zentral), Israel AWS GovCloud (Tel Aviv), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)
2.0.2023 906,0	4,14.322	04. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Kanada (Zentral)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 822.0	4,14.322	30. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Kanada (Zentral)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.20230808,0	4,14.320	24. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Kanada (Zentral)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 727.0	4,14.320	14. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral)

### 6.8.1 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version

sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-amzn-*EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen myapp-component der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als 2.2-amzn-2 aufgeführt.

Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1.4.2	Amazon SageMaker Spark-SDK
emr-ddb	4.16.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	3.2.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.5.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-notebook-env	1.7.0	Conda Env für EMR-Notebooks, das Jupyter Enterprise Gateway enthält
emr-s3-dist-cp	2.22.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	2.1.0	EMR S3Select-Konnektor
emrfs	2,53,0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
flink-client	1.15.1	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
flink-jobmanager-config	1.15.1	Verwaltung von Ressourcen auf EMR-Knoten für Apache JobManager Flink.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	3.2.1-amzn-8.1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	3.2.1-amzn-8.1	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	3.2.1-amzn-8.1	HDFS-Client und -Bibliothek für die Befehlszeile



Komponente	Version	Beschreibung
hadoop-hdfs-namenode	3.2.1-amzn-8.1	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-hdfs-journalnode	3.2.1-amzn-8.1	HDFS-Service zum Verwalten des Hadoop-Dateisystem journals auf HA-Clustern.
hadoop-httpfs-server	3.2.1-amzn-8.1	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	3.2.1-amzn-8.1	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	3.2.1-amzn-8.1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	3.2.1-amzn-8.1	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	3.2.1-amzn-8.1	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	3.2.1-amzn-8.1	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.

Komponente	Version	Beschreibung
hbase-hmaster	2.4.12-amzn-0.1	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	2.4.12-amzn-0.1	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	2.4.12-amzn-0.1	HBase-Befehlszeilen-Client.
hbase-rest-server	2.4.12-amzn-0.1	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	2.4.12-amzn-0.1	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hbase-operator-tools	2.4.12-amzn-0.1	Reparaturtool für Apache-HBase-Cluster.
hcatalog-client	3.1.3-amzn-1.1	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	3.1.3-amzn-1.1	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	3.1.3-amzn-1.1	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	3.1.3-amzn-1.1	Hive-Befehlszeilen-Client.

Komponente	Version	Beschreibung
hive-hbase	3.1.3-amzn-1.1	Hive-hbase client.
hive-metastore-server	3.1.3-amzn-1.1	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	3.1.3-amzn-1.1	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.11.1-amzn-0	Inkrementelles Verarbeitungs-Framework zur Versorgung der Datenpipeline mit geringer Latenz und hoher Effizienz.
hudi-presto	0.11.1-amzn-0	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hudi-trino	0.11.1-amzn-0	Bündel-Bibliothek zum Ausführen von Trino mit Hudi.
hudi-spark	0.11.1-amzn-0	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.
hue-server	4.10.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
iceberg	0.14.0-amzn-0	Apache Iceberg ist ein offenes Tabellenformat für sehr große analytische Datensätze

Komponente	Version	Beschreibung
jupyterhub	1.4.1	Multi-User-Server für Jupyter-Notebooks
livy-server	0.7.1-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP- und Reverse-Proxy-Server.
mxnet	1.9.1	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.68+	MariaDB-Datenbankserver.
nvidia-cuda	11,7,0	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.2.1	Oozie-Befehlszeilen-Client.
oozie-server	5.2.1	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	5.1.2	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-connectors	5.1.2	Apache-Phoenix-Konnektoren für Spark-3

Komponente	Version	Beschreibung
phoenix-query-server	5.1.2	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.273.3-amzn-0	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.273.3-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.273.3-amzn-0	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
trino-coordinator	388-amzn-0	Service zur Annahme von Abfragen und Verwaltung der Abfrageausführung der Trino-Worker.
trino-worker	388-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
trino-client	388-amzn-0	Trino-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Trino-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.

Komponente	Version	Beschreibung
r	4.0.2	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	2.0.0	Apache Ranger Key Management System
spark-client	3.3.0-amzn-0.1	Spark-Befehlszeilen-Clients.
spark-history-server	3.3.0-amzn-0.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	3.3.0-amzn-0.1	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	3.3.0-amzn-0.1	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
spark-rapids	22.06.0-amzn-0	Nvidia-Spark-RAPIDS-Plugin, das Apache Spark mit GPUs beschleunigt.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	2.9.1	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.2	Die Tez-YARN-Anwendung und -Bibliotheken.

Komponente	Version	Beschreibung
webserver	2.4.41+	Apache HTTP-Server.
zeppelin-server	0.10.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.5.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.5.10	ZooKeeper Befehlszeilen-Cli-ent.

## 6.8.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Aktionen zur Neukonfiguration treten auf, wenn Sie eine Konfiguration für Instance-Gruppen in einem laufenden Cluster angeben. Amazon EMR initiiert nur Rekonfigurationsaktionen für die Klassifizierungen, die Sie ändern. Weitere Informationen finden Sie unter [Eine Instance-Gruppe in einem laufenden Cluster neu konfigurieren](#).

### emr-6.8.1-Klassifizierungen

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.	Restarts the Resource Manager service.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
container-executor	Ändern Sie die Werte in der Datei „container-executor.cfg“ Datei von Hadoop YARN.	Not available.
container-log4j	Ändert die Werte in der container-log4j.properties-Datei in Hadoop YARN.	Not available.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Ändern Sie die Docker-bezogenen Einstellungen.	Not available.



Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
emrfs-site	Ändert die EMRFS-Einstellungen.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegistrator, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.	Restarts Flink history server.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.	Restarts Flink history server.
flink-log4j-session	Ändern Sie die Einstellungen von Flink log4j-session.properties für Kubernetes-/Yarn-Sitzungen.	Restarts Flink history server.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.	Restarts Flink history server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.	Not available.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Ändert die Werte in der HBase-Umgebung.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.	Not available.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.	This classification should not be reconfigured.
hdfs-env	Ändert die Werte in der HDFS-Umgebung.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei proto-hive-site .xml von HCatalog.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.	Restarts Hive WebHCat server.
hive	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Hive.	Sets configurations to launch Hive LLAP service.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.	Not available.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.	Not available.
hive-env	Ändert die Werte in der Hive-Umgebung.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.	Not available.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log	Not available.
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.	Not available.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hue-ini	Ändert die Werte in der INI-Datei in Hue.	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.	Restarts Hadoop Httpfs service.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.	Not available.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.	Not available.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Ändern der Werte in der Hudi-Umgebung.	Not available.
hudi-defaults	Ändern Sie die Werte in der hudi-defaults.conf-Datei in Hudi.	Not available.
iceberg-defaults	Ändern Sie die Werte in der iceberg-defaults.conf-Datei von Iceberg.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
jupyter-notebook-conf	Ändert die Werte in der <code>jupyter_notebook_config.py</code> -Datei in Jupyter Notebook.	Not available.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei <code>jupyterhub_config.py</code> .	Not available.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.	Not available.
jupyter-sparkmagic-conf	Ändert die Werte in der <code>config.json</code> -Datei in Sparkmagic.	Not available.
livy-conf	Ändert die Werte in der <code>livy.conf</code> -Datei von Livy.	Restarts Livy Server.
livy-env	Ändert die Werte in der Livy-Umgebung.	Restarts Livy Server.
livy-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Livy.	Restarts Livy Server.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Ändern Sie die Werte in der Datei <code>mapred-site.xml</code> der MapReduce Anwendung.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Ändert die Werte in der Oozie-Umgebung.	Restarts Oozie.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.	Restarts Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.	Not available.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.	Not available.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.	Not available.
pig-env	Ändert die Werte in der Pig-Umgebung.	Not available.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.	Restarts Oozie.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.	Not available.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Ändert die Werte in der config.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)



Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei <code>password-authenticator.properties</code> .	Not available.
presto-env	Ändern Sie die Werte in der <code>presto-env.sh</code> -Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Ändern Sie die Werte in der <code>node.properties</code> -Datei in Presto.	Not available.
presto-connector-blackhole	Ändert die Werte in der <code>blackhole.properties</code> -Datei in Presto.	Not available.
presto-connector-cassandra	Ändert die Werte in der <code>cassandra.properties</code> -Datei in Presto.	Not available.
presto-connector-hive	Ändert die Werte in der <code>hive.properties</code> -Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Ändert die Werte in der <code>jmx.properties</code> -Datei in Presto.	Not available.
presto-connector-kafka	Ändert die Werte in der <code>kafka.properties</code> -Datei in Presto.	Not available.
presto-connector-localfile	Ändert die Werte in der <code>localfile.properties</code> -Datei in Presto.	Not available.
presto-connector-memory	Ändert die Werte in der <code>memory.properties</code> -Datei in Presto.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.	Not available.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.	Not available.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.	Not available.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.	Not available.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.	Not available.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.	Not available.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.	Not available.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.	Not available.
trino-log	Ändern Sie die Werte in der log.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-config	Ändern Sie die Werte in der config.properties-Datei in Trino.	Restarts Trino-Server (for Trino)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-password-authenticator	Ändern Sie die Werte in der Trino-Datei <code>password-authenticator.properties</code> .	Restarts Trino-Server (for Trino)
trino-env	Ändern Sie die Werte in der <code>trino-env.sh</code> -Datei von Trino.	Restarts Trino-Server (for Trino)
trino-node	Ändern Sie die Werte in der <code>node.properties</code> -Datei in Trino.	Not available.
trino-connector-blackhole	Ändern Sie die Werte in der <code>blackhole.properties</code> -Datei in Trino.	Not available.
trino-connector-cassandra	Ändern Sie die Werte in der <code>cassandra.properties</code> -Datei in Trino.	Not available.
trino-connector-hive	Ändern Sie die Werte in der <code>hive.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Ändern Sie die Werte in der <code>iceberg.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Ändern Sie die Werte in der <code>jmx.properties</code> -Datei in Trino.	Not available.
trino-connector-kafka	Ändern Sie die Werte in der <code>kafka.properties</code> -Datei in Trino.	Not available.
trino-connector-localfile	Ändern Sie die Werte in der <code>localfile.properties</code> -Datei in Trino.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-connector-memory	Ändern Sie die Werte in der memory.properties-Datei in Trino.	Not available.
trino-connector-mongodb	Ändern Sie die Werte in der mongodb.properties-Datei in Trino.	Not available.
trino-connector-mysql	Ändern Sie die Werte in der mysql.properties-Datei in Trino.	Not available.
trino-connector-postgresql	Ändern Sie die Werte in der postgresql.properties-Datei in Trino.	Not available.
trino-connector-raptor	Ändern Sie die Werte in der raptor.properties-Datei in Trino.	Not available.
trino-connector-redis	Ändern Sie die Werte in der redis.properties-Datei in Trino.	Not available.
trino-connector-redshift	Ändern Sie die Werte in der redshift.properties-Datei in Trino.	Not available.
trino-connector-tpch	Ändern Sie die Werte in der tpch.properties-Datei in Trino.	Not available.
trino-connector-tpcds	Ändern Sie die Werte in der tpcds.properties-Datei in Trino.	Not available.
ranger-kms-dbks-site	Ändert die Werte in der dbks-site.xml-Datei von Ranger KMS.	Restarts Ranger KMS Server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Ändert die Werte in der Ranger KMS-Umgebung.	Restarts Ranger KMS Server.
ranger-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei von Ranger KMS.	Not available.
ranger-kms-db-ca	Ändert die Werte für die CA-Datei auf S3 für die MySQL SSL-Verbindung mit Ranger KMS.	Not available.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Ändert die Werte in der Spark-Umgebung.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.	Not available.
spark-log4j2	Ändern Sie die Werte in der log4j2.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.	Not available.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.	Not available.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.	Not available.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.	Restart Oozie and HiveServer2.
yarn-env	Ändert die Werte in der YARN-Umgebung.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.	Restarts Zeppelin.
zeppelin-site	Ändern Sie die Konfigurationseinstellungen in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.	Restarts Zookeeper server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.	Restarts Zookeeper server.

## 6.8.1 Änderungsprotokoll

### Änderungsprotokoll für Version 6.8.1 und Versionshinweise

Datum	Ereignis	Beschreibung
2023-08-30	Versionshinweise der Aktualisierung	Den Versionshinweisen wurden mehrere Korrekturen im Zusammenhang mit der Steuerebene hinzugefügt
2023-08-21	Veröffentlichung von Dokumenten	Erste Veröffentlichung der Versionshinweise zu Amazon EMR 6.8.1
2023-08-16	Die Bereitstellung ist abgeschlossen	<a href="#">Amazon EMR 6.8.1 vollständig in allen unterstützten Regionen bereitgestellt</a>
2023-08-04	Erstversion	Amazon EMR 6.8.1 wurde zuerst in begrenzten kommerziellen Regionen bereitgestellt

## Amazon-EMR-Version 6.8.0

### 6.8.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

#### Informationen zur Anwendungsversion

	emr-6.8.0	emr-6.7.0	emr-6.6.0	emr-6.5.0
AWS SDK for Java	1.12.170	1.12,170	1.12.170	1.12,31
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2.12,15	2.12,15	2.12,10	2.12.10
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.15.1	1.14.2	1.14.2	1.14.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4.12	2.4.4	2.4.4	2.4.4
HCatalog	3.1.3	3.1.3	3.1.2	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.3	3.1.3	3.1.2	3.1.2



	emr-6.8.0	emr-6.7.0	emr-6.6.0	emr-6.5.0
Hudi	0.11.1-amzn-0	0.11.0-amzn-0	0.10.1-amzn-0	0.9.0-amzn-1
Hue	4.10.0	4.10.0	4.10.0	4.9.0
Iceberg	0.14.0-amzn-0	0.13.1-amzn-0	0.13.1	0.12.0
JupyterEnterpriseGateway	2.1.0	2.1.0	2.1.0	2.1.0
JupyterHub	1.4.1	1.4.1	1.4.1	1.4.1
Livy	0.7.1	0.7.1	0.7.1	0.7.1
MXNet	1.9.1	1.8.0	1.8.0	1.8.0
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	5.1.2	5.1.2	5.1.2	5.1.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0,273	0,272	0,267	0,261
Spark	3.3.0	3.2.1	3.2.0	3.1.2
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2,9,1	2.4.1	2.4.1	2.4.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	388	378	367	360
Zeppelin	0,10,1	0.10.0	0.10.0	0.10.0
ZooKeeper	3.5.10	3.5.7	3.5.7	3.5.7

## 6.8.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 6.8.0. Änderungen beziehen sich auf Version 6.7.0.

### Neue Features

- Das Amazon-EMR-Schrittfeature unterstützt jetzt Apache.Livy-Endpunkt- und JDBC/ODBC-Clients. Weitere Informationen finden Sie unter [Laufzeit-Rollen für Amazon EMR konfigurieren](#).
- Amazon-EMR-Version 6.8.0 wird mit Apache-HBase-Version 2.4.12 geliefert. Mit dieser HBase-Version können Sie Ihre HBase-Tabellen sowohl archivieren als auch löschen. Der Amazon-S3-Archivierungsprozess benennt alle Tabellendateien in das Archivverzeichnis um. Dies kann ein kostspieliger und langwieriger Prozess sein. Jetzt können Sie den Archivierungsvorgang überspringen und große Tabellen schnell löschen und löschen. Weitere Informationen finden Sie unter [Verwenden der HBase-Shell](#).

### Bekannte Probleme


- Mit Hadoop 3.3.3 wurde eine Änderung in YARN ([YARN-9608](#)) eingeführt, die dafür sorgt, dass Knoten, auf denen Container laufen, außer Betrieb genommen werden, bis die Anwendung abgeschlossen ist. Diese Änderung stellt sicher, dass lokale Daten wie Shuffle-Daten nicht verloren gehen und Sie den Auftrag nicht erneut ausführen müssen. In Amazon EMR 6.8.0 und 6.9.0 kann dieser Ansatz auch zu einer Unterauslastung von Ressourcen auf Clustern mit oder ohne aktivierter verwalteter Skalierung führen.

Mit [Amazon EMR 6.10.0](#) gibt es einen Workaround für dieses Problem, indem Sie den Wert in `yarn-site.xml` von `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` auf `false` setzen. In den Amazon-EMR-Versionen 6.11.0 und höher sowie 6.8.1, 6.9.1 und 6.10.1 ist die Konfiguration standardmäßig auf `false` eingestellt, um dieses Problem zu beheben.

### Änderungen, Verbesserungen und behobene Probleme

- Als Amazon-EMR-Version 6.5.0, 6.6.0 oder 6.7.0 Apache Phoenix-Tabellen über die Apache-Spark-Shell las, erzeugte Amazon EMR eine `NoSuchMethodError`. Amazon EMR Version 6.8.0 behebt dieses Problem.

- Amazon-EMR-Version 6.8.0 enthält [Apache Hudi](#) 0.11.1. Amazon-EMR-6.8.0-Cluster sind jedoch auch mit der Open-Source-Version von Hudi 0.12.0 kompatibel. `hudi-spark3.3-bundle_2.12`
- Amazon-EMR-Version 6.8.0 wird mit Apache Spark 3.3.0 geliefert. Diese Spark-Version verwendet Apache Log4j 2 und die `log4j2.properties`-Datei zur Konfiguration von Log4j in Spark-Prozessen. Wenn Sie Spark im Cluster verwenden oder EMR-Cluster mit benutzerdefinierten Konfigurationsparametern erstellen und ein Upgrade auf Amazon-EMR-Version 6.8.0 durchführen möchten, müssen Sie auf die neue `spark-log4j2`-Konfigurationsklassifizierung und das neue Schlüsselformat für Apache Log4j 2 migrieren. Weitere Informationen finden Sie unter [Migration von Apache Log4j 1.x zu Log4j 2.x](#).
- Wenn Sie einen Cluster mit der neuesten Patch-Version von Amazon EMR 5.36 oder höher, 6.6 oder höher oder 7.0 oder höher starten, verwendet Amazon EMR die neueste Version von Amazon Linux 2023 oder Amazon Linux 2 für das standardmäßige Amazon EMR-AMI. Weitere Informationen finden Sie unter [Verwenden des standardmäßigen Amazon-Linux-AMI für Amazon EMR](#).

 Note

Diese Version erhält keine automatischen AMI-Updates mehr, da sie von einer weiteren Patch-Version abgelöst wurde. Die Patch-Version wird durch die Zahl nach dem zweiten Dezimalpunkt (6.8.**1**) gekennzeichnet. Um zu sehen, ob Sie die neueste Patch-Version verwenden, überprüfen Sie die verfügbaren Versionen im [Versionshandbuch](#) oder überprüfen Sie die Amazon-EMR-Versions-Dropdownliste, wenn Sie einen Cluster in der Konsole erstellen, oder verwenden Sie die [ListReleaseLabels](#)-API- oder [list-release-labels](#)-CLI-Aktion. Um aktuelle Informationen über Neuerscheinungen erhalten, abonnieren Sie den RSS-Feed auf der Seite [Was ist neu?](#).

OsReleaseLabel (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 808.0	4,14.320	24. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			West (Nordkalifornien), USA West (Oregon), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 727.0	4,14.320	14. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 719,0	4,14.320	02. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 628.0	4,14.318	12. Juli 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 612,0	4,14.314	23. Juni 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 504.1	4,14.313	16. Mai 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 418,0	4,14.311	3. Mai 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 404.1	4,14.311	18. April 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)
2.0.2023 404.0	4,14.311	10. April 2023	USA Ost (Nord-Virginia), Europa (Paris)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 320.0	4,14.309	30. März 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 307,0	4,14.305	15. März 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202307,0	4,14.304	22. Februar 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 119.1	4,14.301	3. Februar 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022 210.1	4.14.301	22. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022 103.3	4,14.296	5. Dezember 2022	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022004,0	4,14.294	02. November 2022	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022 912.1	4,14.291	6. September 2022	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

## Bekannte Probleme

- Wenn Sie den DynamoDB-Konnektor mit Spark auf den Amazon-EMR-Versionen 6.6.0, 6.7.0 und 6.8.0 verwenden, geben alle Lesevorgänge aus Ihrer Tabelle ein leeres Ergebnis zurück, obwohl der Eingabe-Split auf nicht leere Daten verweist. Das liegt daran, dass Spark 3.2.0 standardmäßig `spark.hadoopRDD.ignoreEmptySplits` auf `true` einstellt. Um das Problem zu umgehen,

setzen Sie explizit `spark.hadoopRDD.ignoreEmptySplits` auf `false`. Amazon EMR Version 6.9.0 behebt dieses Problem.

- Wenn Sie Spark mit der Formatierung des Hive-Partitionsstandorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den Amazon-EMR-Versionen 5.30.0 bis 5.36.0 und 6.2.0 bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:
  - Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.
  - Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist `s3://bucket/table/p=a` ein Präfix von `s3://bucket/table/p=a b`.
  - Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen UTF-8-Wert, der kleiner als das `/`-Zeichen (U+002F) ist. Beispielsweise fällt das Leerzeichen (U+0020), das in `s3://bucket/table/p=a b` zwischen `a` und `b` vorkommt, in diese Kategorie. Beachten Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: `!"#$%&'()*+,-.` Weitere Informationen finden Sie unter [UTF-8-Kodierungstabelle und Unicode-Zeichen](#).

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.

- Bei den Amazon-EMR-Versionen 5.36.0 und 6.6.0 bis 6.9.0 kann es bei `SecretAgent`- und `RecordServer`-Servicekomponenten aufgrund einer falschen Konfiguration des Dateinamenmusters in den `Log4j2`-Eigenschaften zu einem Verlust von Protokolldaten kommen. Die falsche Konfiguration führt dazu, dass die Komponenten nur eine Protokolldatei pro Tag generieren. Wenn die Rotationsstrategie angewendet wird, überschreibt sie die vorhandene Datei, anstatt wie erwartet eine neue Protokolldatei zu generieren. Um das Problem zu umgehen, verwenden Sie eine Bootstrap-Aktion, um jede Stunde Protokolldateien zu generieren, und fügen Sie eine automatisch inkrementierte Ganzzahl an den Dateinamen an, um die Rotation zu handhaben.

Verwenden Sie für die Versionen Amazon EMR 6.6.0 bis 6.9.0 die folgende Bootstrap-Aktion, wenn Sie einen Cluster starten.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-6x/replace-puppet.sh,Args=[]"
```

Verwenden Sie für Amazon EMR 5.36.0 die folgende Bootstrap-Aktion, wenn Sie einen Cluster starten.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-5x/replace-puppet.sh,Args=[]"
```

Informationen zum Zeitplan der Veröffentlichung finden Sie im [Änderungsprotokoll](#).

## 6.8.0 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.4.2	Amazon SageMaker Spark-SDK
<code>emr-ddb</code>	4.16.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	3.2.0	Praktische Bibliotheken für das Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-kinesis	3.5.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-notebook-env	1.7.0	Conda Env für EMR-Notebooks, das Jupyter Enterprise Gateway enthält
emr-s3-dist-cp	2.22.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	2.1.0	EMR S3Select-Konnektor
emrfs	2,53,0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.15.1	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
flink-jobmanager-config	1.15.1	Verwaltung von Ressourcen auf EMR-Knoten für Apache JobManager Flink.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.

Komponente	Version	Beschreibung
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	3.2.1-amzn-8	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	3.2.1-amzn-8	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	3.2.1-amzn-8	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	3.2.1-amzn-8	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-hdfs-journalnode	3.2.1-amzn-8	HDFS-Service zum Verwalten des Hadoop-Dateisystem journals auf HA-Clustern.
hadoop-httpfs-server	3.2.1-amzn-8	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	3.2.1-amzn-8	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	3.2.1-amzn-8	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.

Komponente	Version	Beschreibung
hadoop-yarn-nodemanager	3.2.1-amzn-8	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	3.2.1-amzn-8	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	3.2.1-amzn-8	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	2.4.12-amzn-0	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	2.4.12-amzn-0	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	2.4.12-amzn-0	HBase-Befehlszeilen-Client.
hbase-rest-server	2.4.12-amzn-0	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	2.4.12-amzn-0	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hbase-operator-tools	2.4.12-amzn-0	Reparaturtool für Apache-HBase-Cluster.



Komponente	Version	Beschreibung
hcatalog-client	3.1.3-amzn-1	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	3.1.3-amzn-1	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	3.1.3-amzn-1	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	3.1.3-amzn-1	Hive-Befehlszeilen-Client.
hive-hbase	3.1.3-amzn-1	Hive-hbase client.
hive-metastore-server	3.1.3-amzn-1	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	3.1.3-amzn-1	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.11.1-amzn-0	Inkrementelles Verarbeitungs-Framework zur Versorgung der Datenpipeline mit geringer Latenz und hoher Effizienz.
hudi-presto	0.11.1-amzn-0	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.

Komponente	Version	Beschreibung
hudi-trino	0.11.1-amzn-0	Bündel-Bibliothek zum Ausführen von Trino mit Hudi.
hudi-spark	0.11.1-amzn-0	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.
hue-server	4.10.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
iceberg	0.14.0-amzn-0	Apache Iceberg ist ein offenes Tabellenformat für sehr große analytische Datensätze
jupyterhub	1.4.1	Multi-User-Server für Jupyter-Notebooks
livy-server	0.7.1-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP- und Reverse-Proxy-Server.
mxnet	1.9.1	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.68+	MariaDB-Datenbankserver.
nvidia-cuda	11,7,0	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.2.1	Oozie-Befehlszeilen-Client.
oozie-server	5.2.1	Service für die Annahme von Oozie Workflow-Anforderungen.

Komponente	Version	Beschreibung
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	5.1.2	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-connectors	5.1.2	Apache-Phoenix-Konnektoren für Spark-3
phoenix-query-server	5.1.2	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.273.3-amzn-0	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.273.3-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.273.3-amzn-0	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
trino-coordinator	388-amzn-0	Service zur Annahme von Abfragen und Verwaltung der Abfrageausführung der Trino-Worker.
trino-worker	388-amzn-0	Service für das Ausführen von Teilen einer Abfrage.

Komponente	Version	Beschreibung
trino-client	388-amzn-0	Trino-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Trino-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	4.0.2	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	2.0.0	Apache Ranger Key Management System
spark-client	3.3.0-amzn-0	Spark-Befehlszeilen-Clients.
spark-history-server	3.3.0-amzn-0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	3.3.0-amzn-0	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	3.3.0-amzn-0	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
spark-rapids	22.06.0-amzn-0	Nvidia-Spark-RAPIDS-Plugin, das Apache Spark mit GPUs beschleunigt.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.

Komponente	Version	Beschreibung
tensorflow	2.9.1	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.2	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.41+	Apache HTTP-Server.
zeppelin-server	0.10.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.5.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.5.10	ZooKeeper Befehlszeilen-Cli-ent.

## 6.8.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Aktionen zur Neukonfiguration treten auf, wenn Sie eine Konfiguration für Instance-Gruppen in einem laufenden Cluster angeben. Amazon EMR initiiert nur Rekonfigurationsaktionen für die Klassifizierungen, die Sie ändern. Weitere Informationen finden Sie unter [Eine Instance-Gruppe in einem laufenden Cluster neu konfigurieren](#).

## emr-6.8.0-Klassifizierungen

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
capacity-scheduler	Ändert die Werte in der capacity-scheduler.xml-Datei in Hadoop.	Restarts the ResourceManager service.
container-executor	Ändern Sie die Werte in der Datei „container-executor.cfg“ Datei von Hadoop YARN.	Not available.
container-log4j	Ändert die Werte in der container-log4j.properties-Datei in Hadoop YARN.	Not available.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Ändern Sie die Docker-bezogenen Einstellungen.	Not available.
emrfs-site	Ändert die EMRFS-Einstellungen.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
		Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegionserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.	Restarts Flink history server.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.	Restarts Flink history server.
flink-log4j-session	Ändern Sie die Einstellungen von Flink log4j-session.properties für Kubernetes-/Yarn-Sitzungen.	Restarts Flink history server.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.	Restarts Flink history server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.	Not available.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.	Not available.



Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Ändert die Werte in der HBase-Umgebung.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.	Not available.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.	This classification should not be reconfigured.
hdfs-env	Ändert die Werte in der HDFS-Umgebung.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei proto-hive-site .xml von HCatalog.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.	Restarts Hive WebHCat server.
hive	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Hive.	Sets configurations to launch Hive LLAP service.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.	Not available.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.	Not available.
hive-env	Ändert die Werte in der Hive-Umgebung.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.	Not available.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log	Not available.
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.	Not available.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hue-ini	Ändert die Werte in der INI-Datei in Hue.	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.	Restarts Hadoop Httpfs service.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.	Not available.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.	Not available.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Ändern der Werte in der Hudi-Umgebung.	Not available.
hudi-defaults	Ändern Sie die Werte in der hudi-defaults.conf-Datei in Hudi.	Not available.
iceberg-defaults	Ändern Sie die Werte in der iceberg-defaults.conf-Datei von Iceberg.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
jupyter-notebook-conf	Ändert die Werte in der <code>jupyter_notebook_config.py</code> -Datei in Jupyter Notebook.	Not available.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei <code>jupyterhub_config.py</code> .	Not available.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.	Not available.
jupyter-sparkmagic-conf	Ändert die Werte in der <code>config.json</code> -Datei in Sparkmagic.	Not available.
livy-conf	Ändert die Werte in der <code>livy.conf</code> -Datei von Livy.	Restarts Livy Server.
livy-env	Ändert die Werte in der Livy-Umgebung.	Restarts Livy Server.
livy-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Livy.	Restarts Livy Server.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Ändern Sie die Werte in der Datei <code>mapred-site.xml</code> der MapReduce Anwendung.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Ändert die Werte in der Oozie-Umgebung.	Restarts Oozie.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.	Restarts Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.	Not available.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.	Not available.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.	Not available.
pig-env	Ändert die Werte in der Pig-Umgebung.	Not available.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.	Restarts Oozie.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.	Not available.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Ändert die Werte in der config.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei <code>password-authenticator.properties</code> .	Not available.
presto-env	Ändern Sie die Werte in der <code>presto-env.sh</code> -Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Ändern Sie die Werte in der <code>node.properties</code> -Datei in Presto.	Not available.
presto-connector-blackhole	Ändert die Werte in der <code>blackhole.properties</code> -Datei in Presto.	Not available.
presto-connector-cassandra	Ändert die Werte in der <code>cassandra.properties</code> -Datei in Presto.	Not available.
presto-connector-hive	Ändert die Werte in der <code>hive.properties</code> -Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Ändert die Werte in der <code>jmx.properties</code> -Datei in Presto.	Not available.
presto-connector-kafka	Ändert die Werte in der <code>kafka.properties</code> -Datei in Presto.	Not available.
presto-connector-localfile	Ändert die Werte in der <code>localfile.properties</code> -Datei in Presto.	Not available.
presto-connector-memory	Ändert die Werte in der <code>memory.properties</code> -Datei in Presto.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-mongodb	Ändert die Werte in der <code>mongodb.properties</code> -Datei in Presto.	Not available.
presto-connector-mysql	Ändert die Werte in der <code>mysql.properties</code> -Datei in Presto.	Not available.
presto-connector-postgresql	Ändert die Werte in der <code>postgresql.properties</code> -Datei in Presto.	Not available.
presto-connector-raptor	Ändert die Werte in der <code>raptor.properties</code> -Datei in Presto.	Not available.
presto-connector-redis	Ändert die Werte in der <code>redis.properties</code> -Datei in Presto.	Not available.
presto-connector-redshift	Ändert die Werte in der <code>redshift.properties</code> -Datei.	Not available.
presto-connector-tpch	Ändert die Werte in der <code>tpch.properties</code> -Datei in Presto.	Not available.
presto-connector-tpcds	Ändert die Werte in der <code>tpcds.properties</code> -Datei in Presto.	Not available.
trino-log	Ändern Sie die Werte in der <code>log.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)
trino-config	Ändern Sie die Werte in der <code>config.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)



Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-password-authenticator	Ändern Sie die Werte in der Trino-Datei <code>password-authenticator.properties</code> .	Restarts Trino-Server (for Trino)
trino-env	Ändern Sie die Werte in der <code>trino-env.sh</code> -Datei von Trino.	Restarts Trino-Server (for Trino)
trino-node	Ändern Sie die Werte in der <code>node.properties</code> -Datei in Trino.	Not available.
trino-connector-blackhole	Ändern Sie die Werte in der <code>blackhole.properties</code> -Datei in Trino.	Not available.
trino-connector-cassandra	Ändern Sie die Werte in der <code>cassandra.properties</code> -Datei in Trino.	Not available.
trino-connector-hive	Ändern Sie die Werte in der <code>hive.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Ändern Sie die Werte in der <code>iceberg.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Ändern Sie die Werte in der <code>jmx.properties</code> -Datei in Trino.	Not available.
trino-connector-kafka	Ändern Sie die Werte in der <code>kafka.properties</code> -Datei in Trino.	Not available.
trino-connector-localfile	Ändern Sie die Werte in der <code>localfile.properties</code> -Datei in Trino.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-connector-memory	Ändern Sie die Werte in der memory.properties-Datei in Trino.	Not available.
trino-connector-mongodb	Ändern Sie die Werte in der mongodb.properties-Datei in Trino.	Not available.
trino-connector-mysql	Ändern Sie die Werte in der mysql.properties-Datei in Trino.	Not available.
trino-connector-postgresql	Ändern Sie die Werte in der postgresql.properties-Datei in Trino.	Not available.
trino-connector-raptor	Ändern Sie die Werte in der raptor.properties-Datei in Trino.	Not available.
trino-connector-redis	Ändern Sie die Werte in der redis.properties-Datei in Trino.	Not available.
trino-connector-redshift	Ändern Sie die Werte in der redshift.properties-Datei in Trino.	Not available.
trino-connector-tpch	Ändern Sie die Werte in der tpch.properties-Datei in Trino.	Not available.
trino-connector-tpcds	Ändern Sie die Werte in der tpcds.properties-Datei in Trino.	Not available.
ranger-kms-dbks-site	Ändert die Werte in der dbks-site.xml-Datei von Ranger KMS.	Restarts Ranger KMS Server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Ändert die Werte in der Ranger KMS-Umgebung.	Restarts Ranger KMS Server.
ranger-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei von Ranger KMS.	Not available.
ranger-kms-db-ca	Ändert die Werte für die CA-Datei auf S3 für die MySQL SSL-Verbindung mit Ranger KMS.	Not available.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Ändert die Werte in der Spark-Umgebung.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.	Not available.
spark-log4j2	Ändern Sie die Werte in der log4j2.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.	Not available.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.	Not available.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.	Not available.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.	Restart Oozie and HiveServer2.
yarn-env	Ändert die Werte in der YARN-Umgebung.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.	Restarts Zeppelin.
zeppelin-site	Ändern Sie die Konfigurationseinstellungen in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.	Restarts Zookeeper server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.	Restarts Zookeeper server.

## 6.8.0 Änderungsprotokoll

### Änderungsprotokoll für Version 6.8.0 und Versionshinweise

Datum	Ereignis	Beschreibung
2023-08-21	Aktualisierung	Es wurde ein bekanntes Problem mit Hadoop 3.3.3 hinzugefügt.
2023-07-26	Aktualisierung	Neue Betriebssystem-Release-Labels 2.0.20230612.0 und 2.0.20230628.0 .
06.09.2022	Die Bereitstellung ist abgeschlossen	Amazon EMR 6.8 wurde vollständig in allen <a href="#">unterstützten</a> Regionen bereitgestellt
06.09.2022	Erste Veröffentlichung	Erste Veröffentlichung der Versionshinweise zu Amazon EMR 6.8
31.08.2022	Erstversion	Amazon EMR 6.8 wurde zuerst in begrenzten kommerziellen Regionen veröffentlicht

## Amazon-EMR-Version 6.7.0

### 6.7.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Liviy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

#### Informationen zur Anwendungsversion

	emr-6.7.0	emr-6.6.0	emr-6.5.0	emr-6.4.0
AWS SDK for Java	1.12.170	1.12.170	1.12,31	1.12.31
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2.12,15	2.12,10	2.12.10	2.12.10
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.14.2	1.14.2	1.14.0	1.13.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2

	emr-6.7.0	emr-6.6.0	emr-6.5.0	emr-6.4.0
HBase	2.4.4	2.4.4	2.4.4	2.4.4
HCatalog	3.1.3	3.1.2	3.1.2	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.3	3.1.2	3.1.2	3.1.2
Hudi	0.11.0-amzn-0	0.10.1-amzn-0	0.9.0-amzn-1	0.8.0-amzn-0
Hue	4.10.0	4.10.0	4.9.0	4.9.0
Iceberg	0.13.1-amzn-0	0.13.1	0.12.0	-
JupyterEnterpriseGateway	2.1.0	2.1.0	2.1.0	2.1.0
JupyterHub	1.4.1	1.4.1	1.4.1	1.4.1
Livy	0.7.1	0.7.1	0.7.1	0.7.1
MXNet	1.8.0	1.8.0	1.8.0	1.8.0
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	5.1.2	5.1.2	5.1.2	5.1.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0,272	0,267	0,261	0,254,1
Spark	3.2.1	3.2.0	3.1.2	3.1.2
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.4.1	2.4.1	2.4.1	2.4.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2

	emr-6.7.0	emr-6.6.0	emr-6.5.0	emr-6.4.0
Trino (PrestoSQL)	378	367	360	359
Zeppelin	0.10.0	0.10.0	0.10.0	0.9.0
ZooKeeper	3,5.7	3.5.7	3.5.7	3.5.7

## 6.7.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 6.7.0. Änderungen beziehen sich auf Version 6.6.0.

Erste Version: 15. Juli 2022

### Neue Features

- Amazon EMR unterstützt jetzt Apache Spark 3.2.1, Apache Hive 3.1.3, HUDI 0.11, PrestoDB 0.272 und Trino 0.378.
- Unterstützt Zugriffskontrollen auf Basis der IAM-Rolle und Lake Formation mit EMR-Schritten (Spark, Hive) für Amazon EMR auf EC2-Clustern.
- Unterstützt Apache-Spark-Datendefinitionsanweisungen auf Apache-Ranger-fähigen Clustern. Dies beinhaltet jetzt die Unterstützung für Trino-Anwendungen, die Apache-Hive-Metadaten auf Apache-Ranger-fähigen Clustern lesen und schreiben. Weitere Informationen finden Sie unter [Aktivieren von Verbund-Governance mit Trino und Apache Ranger auf Amazon EMR](#).
- Wenn Sie einen Cluster mit der neuesten Patch-Version von Amazon EMR 5.36 oder höher, 6.6 oder höher oder 7.0 oder höher starten, verwendet Amazon EMR die neueste Version von Amazon Linux 2023 oder Amazon Linux 2 für das standardmäßige Amazon EMR-AMI. Weitere Informationen finden Sie unter [Verwenden des standardmäßigen Amazon-Linux-AMI für Amazon EMR](#).



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 223.0	4,14.336	8. März 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)
2.0.2024131,0	4,14.336	14. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral), AWS GovCloud (US-West), AWS GovCloud (US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 124,0	4,14.336	7. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral), AWS GovCloud (US-West), AWS GovCloud (US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024109,0	4,14.334	24. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral), AWS GovCloud (US-West), AWS GovCloud (US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 218,0	4,14.330	2. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral), AWS GovCloud (US-West), AWS GovCloud (US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202306,0	4,14.330	22. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada AWS GovCloud (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023116,0	4,14.328	11. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada AWS GovCloud (Zentral), AWS GovCloud (US-West),



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023101,0	4,14.327	16. November 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada AWS GovCloud (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023020,1	4,14.326	7. November 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada AWS GovCloud (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023012,1	4,14.326	26. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada AWS GovCloud (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 926,0	4,14.322	19. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada AWS GovCloud (Zentral), AWS GovCloud (US-West),



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)
2.0.2023 906,0	4,14.322	04. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 822.0	4,14.322	30. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.20230808.0	4,14.320	24. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 727.0	4,14.320	14. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023719,0	4,14.320	02. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 628.0	4,14.318	12. Juli 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 612,0	4,14.314	23. Juni 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 504.1	4,14.313	16. Mai 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 418,0	4,14.311	3. Mai 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 404.1	4,14.311	18. April 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)
2.0.2023 404.0	4,14.311	10. April 2023	USA Ost (Nord-Virginia), Europa (Paris)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 320.0	4,14.309	30. März 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202307,0	4,14.305	15. März 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202307,0	4,14.304	22. Februar 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023119.1	4,14.301	3. Februar 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022 4.14.301 210.1		22. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022103.3	4,14.296	5. Dezember 2022	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022004,0	4,14.294	02. November 2022	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022 912.1	4,14.291	7. Oktober 2022	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)
2.0.2022 719.0	4,14.287	10. August 2022	us-west-1 , eu-west-3 , eu-north-1 , ap-south-1 , me-south-1

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022.606.1	4,14.281	15. Juli 2022	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

## Bekannte Probleme

- Wenn Amazon-EMR-Version 6.5.0, 6.6.0 oder 6.7.0 Apache Phoenix-Tabellen über die Apache Spark-Shell liest, tritt `NoSuchMethodError` ein, weil Amazon EMR ein falsches `Hbase.compat.version` verwendet. Amazon EMR Version 6.8.0 behebt dieses Problem.
- Wenn Sie den DynamoDB-Konnektor mit Spark auf den Amazon-EMR-Versionen 6.6.0, 6.7.0 und 6.8.0 verwenden, geben alle Lesevorgänge aus Ihrer Tabelle ein leeres Ergebnis zurück, obwohl

der Eingabe-Split auf nicht leere Daten verweist. Das liegt daran, dass Spark 3.2.0 standardmäßig `spark.hadoopRDD.ignoreEmptySplits` auf `true` einstellt. Um das Problem zu umgehen, setzen Sie explizit `spark.hadoopRDD.ignoreEmptySplits` auf `false`. Amazon EMR Version 6.9.0 behebt dieses Problem.

- Wenn Sie Spark mit der Formatierung des Hive-Partitionsstandorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den Amazon-EMR-Versionen 5.30.0 bis 5.36.0 und 6.2.0 bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:
  - Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.
  - Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist `s3://bucket/table/p=a` ein Präfix von `s3://bucket/table/p=a b`.
  - Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen UTF-8-Wert, der kleiner als das `/`-Zeichen (U+002F) ist. Beispielsweise fällt das Leerzeichen (U+0020), das in `s3://bucket/table/p=a b` zwischen `a` und `b` vorkommt, in diese Kategorie. Beachten Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: `!"#$%&'()*+,-.` Weitere Informationen finden Sie unter [UTF-8-Kodierungstabelle und Unicode-Zeichen](#).

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.

- Bei den Amazon-EMR-Versionen 5.36.0 und 6.6.0 bis 6.9.0 kann es bei `SecretAgent`- und `RecordServer`-Servicekomponenten aufgrund einer falschen Konfiguration des Dateinamenmusters in den `Log4j2`-Eigenschaften zu einem Verlust von Protokolldaten kommen. Die falsche Konfiguration führt dazu, dass die Komponenten nur eine Protokolldatei pro Tag generieren. Wenn die Rotationsstrategie angewendet wird, überschreibt sie die vorhandene Datei, anstatt wie erwartet eine neue Protokolldatei zu generieren. Um das Problem zu umgehen, verwenden Sie eine Bootstrap-Aktion, um jede Stunde Protokolldateien zu generieren, und fügen Sie eine automatisch inkrementierte Ganzzahl an den Dateinamen an, um die Rotation zu handhaben.

Verwenden Sie für die Versionen Amazon EMR 6.6.0 bis 6.9.0 die folgende Bootstrap-Aktion, wenn Sie einen Cluster starten.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-6x/replace-puppet.sh,Args=[]"
```

Verwenden Sie für Amazon EMR 5.36.0 die folgende Bootstrap-Aktion, wenn Sie einen Cluster starten.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-5x/replace-puppet.sh,Args=[]"
```

- Die `GetClusterSessionCredentials` API wird nicht mit Clustern unterstützt, die in Amazon EMR 6.7 oder niedriger ausgeführt werden.

## 6.7.0 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.4.1	Amazon SageMaker Spark-SDK
<code>emr-ddb</code>	4.16.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	3.2.0	Praktische Bibliotheken für das Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-kinesis	3.5.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-notebook-env	1.6.0	Conda Env für EMR-Notebooks, das Jupyter Enterprise Gateway enthält
emr-s3-dist-cp	2.22.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	2.1.0	EMR S3Select-Konnektor
emrfs	2,52,0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1,14,2	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
flink-jobmanager-config	1.14.2	Verwaltung von Ressourcen auf EMR-Knoten für Apache JobManager Flink.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.

Komponente	Version	Beschreibung
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	3.2.1-amzn-7	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	3.2.1-amzn-7	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	3.2.1-amzn-7	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	3.2.1-amzn-7	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-hdfs-journalnode	3.2.1-amzn-7	HDFS-Service zum Verwalten des Hadoop-Dateisystem journals auf HA-Clustern.
hadoop-httpfs-server	3.2.1-amzn-7	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	3.2.1-amzn-7	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	3.2.1-amzn-7	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.

Komponente	Version	Beschreibung
hadoop-yarn-nodemanager	3.2.1-amzn-7	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	3.2.1-amzn-7	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	3.2.1-amzn-7	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	2.4.4-amzn-3	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	2.4.4-amzn-3	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	2.4.4-amzn-3	HBase-Befehlszeilen-Client.
hbase-rest-server	2.4.4-amzn-3	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	2.4.4-amzn-3	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hbase-operator-tools	2.4.4-amzn-3	Reparaturtool für Apache-HBase-Cluster.



Komponente	Version	Beschreibung
hcatalog-client	3.1.3-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	3.1.3-amzn-0	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	3.1.3-amzn-0	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	3.1.3-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	3.1.3-amzn-0	Hive-hbase client.
hive-metastore-server	3.1.3-amzn-0	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	3.1.3-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.11.0-amzn-0	Inkrementelles Verarbeitungs-Framework zur Versorgung der Datenpipeline mit geringer Latenz und hoher Effizienz.
hudi-presto	0.11.0-amzn-0	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.

Komponente	Version	Beschreibung
hudi-trino	0.11.0-amzn-0	Bündel-Bibliothek zum Ausführen von Trino mit Hudi.
hudi-spark	0.11.0-amzn-0	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.
hue-server	4.10.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
iceberg	0.13.1-amzn-0	Apache Iceberg ist ein offenes Tabellenformat für sehr große analytische Datensätze
jupyterhub	1.4.1	Multi-User-Server für Jupyter-Notebooks
livy-server	0.7.1-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP- und Reverse-Proxy-Server.
mxnet	1.8.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.68+	MariaDB-Datenbankserver.
nvidia-cuda	11,0,194	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.2.1	Oozie-Befehlszeilen-Client.
oozie-server	5.2.1	Service für die Annahme von Oozie Workflow-Anforderungen.

Komponente	Version	Beschreibung
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	5.1.2	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-connectors	5.1.2	Apache-Phoenix-Konnektoren für Spark-3
phoenix-query-server	5.1.2	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.272-amzn-0	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.272-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.272-amzn-0	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
trino-coordinator	378-amzn-0	Service zur Annahme von Abfragen und Verwaltung der Abfrageausführung der Trino-Worker.
trino-worker	378-amzn-0	Service für das Ausführen von Teilen einer Abfrage.

Komponente	Version	Beschreibung
trino-client	378-amzn-0	Trino-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Trino-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	4.0.2	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	2.0.0	Apache Ranger Key Management System
spark-client	3.2.1-amzn-0	Spark-Befehlszeilen-Clients.
spark-history-server	3.2.1-amzn-0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	3.2.1-amzn-0	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	3.2.1-amzn-0	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
spark-rapids	22.02.0-amzn-1	Nvidia-Spark-RAPIDS-Plugin, das Apache Spark mit GPUs beschleunigt.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.

Komponente	Version	Beschreibung
tensorflow	2.4.1	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.2	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.41+	Apache HTTP-Server.
zeppelin-server	0.10.0	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.5.7	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.5.7	ZooKeeper Befehlszeilen-Cli-ent.

## 6.7.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Aktionen zur Neukonfiguration treten auf, wenn Sie eine Konfiguration für Instance-Gruppen in einem laufenden Cluster angeben. Amazon EMR initiiert nur Rekonfigurationsaktionen für die Klassifizierungen, die Sie ändern. Weitere Informationen finden Sie unter [Eine Instance-Gruppe in einem laufenden Cluster neu konfigurieren](#).

## emr-6.7.0-Klassifizierungen

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
capacity-scheduler	Ändert die Werte in der capacity-scheduler.xml-Datei in Hadoop.	Restarts the ResourceManager service.
container-executor	Ändern Sie die Werte in der Datei „container-executor.cfg“ Datei von Hadoop YARN.	Not available.
container-log4j	Ändert die Werte in der container-log4j.properties-Datei in Hadoop YARN.	Not available.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Ändern Sie die Docker-bezogenen Einstellungen.	Not available.
emrfs-site	Ändert die EMRFS-Einstellungen.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
		Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegionserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.	Restarts Flink history server.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.	Restarts Flink history server.
flink-log4j-session	Ändern Sie die Einstellungen von Flink log4j-session.properties für Kubernetes-/Yarn-Sitzungen.	Restarts Flink history server.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.	Restarts Flink history server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.	Not available.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.	Not available.



Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Ändert die Werte in der HBase-Umgebung.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.	Not available.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.	This classification should not be reconfigured.
hdfs-env	Ändert die Werte in der HDFS-Umgebung.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei proto-hive-site .xml von HCatalog.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.	Restarts Hive WebHCat server.
hive	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Hive.	Sets configurations to launch Hive LLAP service.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.	Not available.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.	Not available.
hive-env	Ändert die Werte in der Hive-Umgebung.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.	Not available.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log	Not available.
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.	Not available.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hue-ini	Ändert die Werte in der INI-Datei in Hue.	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.	Restarts Hadoop Httpfs service.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.	Not available.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.	Not available.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Ändern der Werte in der Hudi-Umgebung.	Not available.
hudi-defaults	Ändern Sie die Werte in der hudi-defaults.conf-Datei in Hudi.	Not available.
iceberg-defaults	Ändern Sie die Werte in der iceberg-defaults.conf-Datei von Iceberg.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
jupyter-notebook-conf	Ändert die Werte in der <code>jupyter_notebook_config.py</code> -Datei in Jupyter Notebook.	Not available.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei <code>jupyterhub_config.py</code> .	Not available.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.	Not available.
jupyter-sparkmagic-conf	Ändert die Werte in der <code>config.json</code> -Datei in Sparkmagic.	Not available.
livy-conf	Ändert die Werte in der <code>livy.conf</code> -Datei von Livy.	Restarts Livy Server.
livy-env	Ändert die Werte in der Livy-Umgebung.	Restarts Livy Server.
livy-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Livy.	Restarts Livy Server.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Ändern Sie die Werte in der Datei <code>mapred-site.xml</code> der MapReduce Anwendung.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Ändert die Werte in der Oozie-Umgebung.	Restarts Oozie.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.	Restarts Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.	Not available.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.	Not available.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.	Not available.
pig-env	Ändert die Werte in der Pig-Umgebung.	Not available.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.	Restarts Oozie.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.	Not available.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Ändert die Werte in der config.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei <code>password-authenticator.properties</code> .	Not available.
presto-env	Ändern Sie die Werte in der <code>presto-env.sh</code> -Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Ändern Sie die Werte in der <code>node.properties</code> -Datei in Presto.	Not available.
presto-connector-blackhole	Ändert die Werte in der <code>blackhole.properties</code> -Datei in Presto.	Not available.
presto-connector-cassandra	Ändert die Werte in der <code>cassandra.properties</code> -Datei in Presto.	Not available.
presto-connector-hive	Ändert die Werte in der <code>hive.properties</code> -Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Ändert die Werte in der <code>jmx.properties</code> -Datei in Presto.	Not available.
presto-connector-kafka	Ändert die Werte in der <code>kafka.properties</code> -Datei in Presto.	Not available.
presto-connector-localfile	Ändert die Werte in der <code>localfile.properties</code> -Datei in Presto.	Not available.
presto-connector-memory	Ändert die Werte in der <code>memory.properties</code> -Datei in Presto.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-mongodb	Ändert die Werte in der <code>mongodb.properties</code> -Datei in Presto.	Not available.
presto-connector-mysql	Ändert die Werte in der <code>mysql.properties</code> -Datei in Presto.	Not available.
presto-connector-postgresql	Ändert die Werte in der <code>postgresql.properties</code> -Datei in Presto.	Not available.
presto-connector-raptor	Ändert die Werte in der <code>raptor.properties</code> -Datei in Presto.	Not available.
presto-connector-redis	Ändert die Werte in der <code>redis.properties</code> -Datei in Presto.	Not available.
presto-connector-redshift	Ändert die Werte in der <code>redshift.properties</code> -Datei.	Not available.
presto-connector-tpch	Ändert die Werte in der <code>tpch.properties</code> -Datei in Presto.	Not available.
presto-connector-tpcds	Ändert die Werte in der <code>tpcds.properties</code> -Datei in Presto.	Not available.
trino-log	Ändern Sie die Werte in der <code>log.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)
trino-config	Ändern Sie die Werte in der <code>config.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)



Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-password-authenticator	Ändern Sie die Werte in der Trino-Datei <code>password-authenticator.properties</code> .	Restarts Trino-Server (for Trino)
trino-env	Ändern Sie die Werte in der <code>trino-env.sh</code> -Datei von Trino.	Restarts Trino-Server (for Trino)
trino-node	Ändern Sie die Werte in der <code>node.properties</code> -Datei in Trino.	Not available.
trino-connector-blackhole	Ändern Sie die Werte in der <code>blackhole.properties</code> -Datei in Trino.	Not available.
trino-connector-cassandra	Ändern Sie die Werte in der <code>cassandra.properties</code> -Datei in Trino.	Not available.
trino-connector-hive	Ändern Sie die Werte in der <code>hive.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Ändern Sie die Werte in der <code>iceberg.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Ändern Sie die Werte in der <code>jmx.properties</code> -Datei in Trino.	Not available.
trino-connector-kafka	Ändern Sie die Werte in der <code>kafka.properties</code> -Datei in Trino.	Not available.
trino-connector-localfile	Ändern Sie die Werte in der <code>localfile.properties</code> -Datei in Trino.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-connector-memory	Ändern Sie die Werte in der memory.properties-Datei in Trino.	Not available.
trino-connector-mongodb	Ändern Sie die Werte in der mongodb.properties-Datei in Trino.	Not available.
trino-connector-mysql	Ändern Sie die Werte in der mysql.properties-Datei in Trino.	Not available.
trino-connector-postgresql	Ändern Sie die Werte in der postgresql.properties-Datei in Trino.	Not available.
trino-connector-raptor	Ändern Sie die Werte in der raptor.properties-Datei in Trino.	Not available.
trino-connector-redis	Ändern Sie die Werte in der redis.properties-Datei in Trino.	Not available.
trino-connector-redshift	Ändern Sie die Werte in der redshift.properties-Datei in Trino.	Not available.
trino-connector-tpch	Ändern Sie die Werte in der tpch.properties-Datei in Trino.	Not available.
trino-connector-tpcds	Ändern Sie die Werte in der tpcds.properties-Datei in Trino.	Not available.
ranger-kms-dbks-site	Ändert die Werte in der dbks-site.xml-Datei von Ranger KMS.	Restarts Ranger KMS Server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Ändert die Werte in der Ranger KMS-Umgebung.	Restarts Ranger KMS Server.
ranger-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei von Ranger KMS.	Not available.
ranger-kms-db-ca	Ändert die Werte für die CA-Datei auf S3 für die MySQL SSL-Verbindung mit Ranger KMS.	Not available.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Ändert die Werte in der Spark-Umgebung.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.	Not available.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.	Not available.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.	Not available.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.	Not available.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.	Restart Oozie and HiveServer2.
yarn-env	Ändert die Werte in der YARN-Umgebung.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.	Restarts Zeppelin.
zeppelin-site	Ändern Sie die Konfigurationseinstellungen in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.	Restarts Zookeeper server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.	Restarts Zookeeper server.

## Amazon-EMR-Version 6.6.0

### 6.6.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

### Informationen zur Anwendungsversion

	emr-6.6.0	emr-6.5.0	emr-6.4.0	emr-6.3.1
AWS SDK for Java	1.12.170	1.12,31	1.12.31	1.11.977
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2.12,10	2.12.10	2.12.10	2.12.10

	emr-6.6.0	emr-6.5.0	emr-6.4.0	emr-6.3.1
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.14.2	1.14.0	1.13.1	1.12.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4.4	2.4.4	2.4.4	2.2.6
HCatalog	3.1.2	3.1.2	3.1.2	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.2	3.1.2	3.1.2	3.1.2
Hudi	0.10.1-amzn-0	0.9.0-amzn-1	0.8.0-amzn-0	0.7.0-amzn-0
Hue	4.10.0	4.9.0	4.9.0	4.9.0
Iceberg	0.13.1	0.12.0	-	-
JupyterEnterpriseGateway	2.1.0	2.1.0	2.1.0	2.1.0
JupyterHub	1.4.1	1.4.1	1.4.1	1.2.2
Livy	0.7.1	0.7.1	0.7.1	0.7.0
MXNet	1.8.0	1.8.0	1.8.0	1.7.0
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	5.1.2	5.1.2	5.1.2	5.0.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0

	emr-6.6.0	emr-6.5.0	emr-6.4.0	emr-6.3.1
Presto	0,267	0,261	0,254,1	0,245,1
Spark	3.2.0	3.1.2	3.1.2	3.1.1
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.4.1	2.4.1	2.4.1	2.4.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	367	360	359	350
Zeppelin	0.10.0	0.10.0	0.9.0	0.9.0
ZooKeeper	3,5.7	3.5.7	3.5.7	3.4.14

## 6.6.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 6.6.0. Änderungen beziehen sich auf Version 6.5.0.

Erste Version: 09. Mai 2022

Letzte Aktualisierung der Dokumentation: 15. Juni 2022

### Neue Features

- Amazon EMR 6.6 unterstützt jetzt Apache Spark 3.2, Apache Spark RAPIDS 22.02, CUDA 11, Apache Hudi 0.10.1, Apache Iceberg 0.13, Trino 0.367 und PrestoDB 0.267.
- Wenn Sie einen Cluster mit der neuesten Patch-Version von Amazon EMR 5.36 oder höher, 6.6 oder höher oder 7.0 oder höher starten, verwendet Amazon EMR die neueste Version von Amazon Linux 2023 oder Amazon Linux 2 für das standardmäßige Amazon EMR-AMI. Weitere Informationen finden Sie unter [Verwenden des standardmäßigen Amazon-Linux-AMI für Amazon EMR](#).

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 223.0	4,14.336	8. März 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West),



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)
2.0.2024 131,0	4,14.336	14. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral), AWS GovCloud (US-West), AWS GovCloud (US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 124,0	4,14.336	7. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral), AWS GovCloud (US-West), AWS GovCloud (US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024109,0	4,14.334	24. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral), AWS GovCloud (US-West), AWS GovCloud (US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 218,0	4,14.330	2. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral), AWS GovCloud (US-West), AWS GovCloud (US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202306,0	4,14.330	22. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada AWS GovCloud (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023116,0	4,14.328	11. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada AWS GovCloud (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023101,0	4,14.327	16. November 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada AWS GovCloud (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023020,1	4,14.326	7. November 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada AWS GovCloud (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023012,1	4,14.326	26. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada AWS GovCloud (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 926,0	4,14.322	19. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada AWS GovCloud (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)
2.0.2023 906,0	4,14.322	04. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral)



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 822.0	4,14.322	30. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.20230808,0	4,14.320	24. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 727.0	4,14.320	14. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 719,0	4,14.320	02. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 628.0	4,14.318	12. Juli 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 612,0	4,14.314	23. Juni 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 504.1	4,14.313	16. Mai 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 418,0	4,14.311	3. Mai 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 404.1	4,14.311	18. April 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)
2.0.2023 404.0	4,14.311	10. April 2023	USA Ost (Nord-Virginia), Europa (Paris)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 320.0	4,14.309	30. März 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202307,0	4,14.305	15. März 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202307,0	4,14.304	22. Februar 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023119.1	4,14.301	3. Februar 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022 4.14.301 210.1		22. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022103.3	4,14.296	5. Dezember 2022	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022004,0	4,14.294	02. November 2022	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022 912.1	4,14.291	7. Oktober 2022	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)
2.0.2022 805.0	4,14.287	30. August 2022	us-west-1

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022 719.0	4,14.287	10. August 2022	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022 426,0	4,14.281	10. Juni 2022	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022.406.1	4,14.275	2. Mai 2022	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

- Mit Amazon EMR 6.6 und höher werden Anwendungen, die Log4j 1.x und Log4j 2.x verwenden, auf Log4j 1.2.17 (oder höher) bzw. Log4j 2.17.1 (oder höher) aktualisiert, sodass die [Bootstrap-Aktionen](#) die zur Behebung der CVE-Probleme bereitgestellt werden, nicht erforderlich sind.
- [Verwaltete Skalierung] Spark Shuffle Data Managed Scaling Optimization – Für Amazon-EMR-Versionen 5.34.0 und höher sowie EMR-Versionen 6.4.0 und höher unterstützt verwaltete Skalierung jetzt Spark Shuffle Data (Daten, die Spark partitionsübergreifend verteilt, um bestimmte Operationen auszuführen). Weitere Informationen zu Shuffle-Vorgängen finden Sie unter [Nutzung](#)

[von Managed Scaling in Amazon EMR](#) im Verwaltungshandbuch für Amazon EMR und [Spark-Programmierungs-Handbuch](#).

- Ab Amazon EMR 5.32.0 und 6.5.0 ist die dynamische Executor-Größenanpassung für Apache Spark standardmäßig aktiviert. Sie können dieses Feature ein- oder ausschalten, indem Sie den Konfigurationsparameter `spark.yarn.heterogeneousExecutors.enabled` verwenden.

### Änderungen, Verbesserungen und behobene Probleme

- Amazon EMR reduziert die Cluster-Startup-Zeit für Cluster, die die EMR-Standard-AMI-Option verwenden und nur gängige Anwendungen wie Apache Hadoop, Apache Spark und Apache Hive installieren, um durchschnittlich bis zu 80 Sekunden.

### Bekannte Probleme

- Wenn Amazon-EMR-Version 6.5.0, 6.6.0 oder 6.7.0 Apache Phoenix-Tabellen über die Apache Spark-Shell liest, tritt `NoSuchMethodError` ein, weil Amazon EMR ein falsches `Hbase.compat.version` verwendet. Amazon EMR Version 6.8.0 behebt dieses Problem.
- Wenn Sie den DynamoDB-Konnektor mit Spark auf den Amazon-EMR-Versionen 6.6.0, 6.7.0 und 6.8.0 verwenden, geben alle Lesevorgänge aus Ihrer Tabelle ein leeres Ergebnis zurück, obwohl der Eingabe-Split auf nicht leere Daten verweist. Das liegt daran, dass Spark 3.2.0 standardmäßig `spark.hadoopRDD.ignoreEmptySplits` auf `true` einstellt. Um das Problem zu umgehen, setzen Sie explizit `spark.hadoopRDD.ignoreEmptySplits` auf `false`. Amazon EMR Version 6.9.0 behebt dieses Problem.
- Auf Trino-Clustern mit langer Laufzeit aktiviert Amazon EMR 6.6.0 die Garbage-Collection-Protokollierungs-Parameter in der Trino `jvm.config`, um bessere Einblicke in die Garbage-Collection-Protokolle zu erhalten. Durch diese Änderung werden viele Garbage-Collection-Protokolle an die Datei `launcher.log` (`/var/log/trino/launcher.log`) angehängt. Wenn Sie Trino-Cluster in Amazon EMR 6.6.0 ausführen, kann es vorkommen, dass Knoten aufgrund der angehängten Protokolle nicht mehr genügend Festplattenspeicher haben, nachdem der Cluster einige Tage lang ausgeführt wurde.

Um dieses Problem zu umgehen, führen Sie das folgende Skript als Bootstrap-Aktion aus, um die Garbage-Collection-Protokollierungs-Parameter in `jvm.config` zu deaktivieren, während Sie den Cluster für Amazon EMR 6.6.0 erstellen oder klonen.

```
#!/bin/bash
```

```
set -ex
PRESTO_PUPPET_DIR='/var/aws/emr/bigtop-deploy/puppet/modules/trino'
sudo bash -c "sed -i '/-Xlog/d' ${PRESTO_PUPPET_DIR}/templates/jvm.config"
```

- Wenn Sie Spark mit der Formatierung des Hive-Partitionsstandorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den Amazon-EMR-Versionen 5.30.0 bis 5.36.0 und 6.2.0 bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:
  - Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.
  - Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist `s3://bucket/table/p=a` ein Präfix von `s3://bucket/table/p=a b`.
  - Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen UTF-8-Wert, der kleiner als das `/`-Zeichen (U+002F) ist. Beispielsweise fällt das Leerzeichen (U+0020), das in `s3://bucket/table/p=a b` zwischen `a` und `b` vorkommt, in diese Kategorie. Beachten Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: `!"#$%&'()*+,-.` Weitere Informationen finden Sie unter [UTF-8-Kodierungstabelle und Unicode-Zeichen](#).

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.

- Bei den Amazon-EMR-Versionen 5.36.0 und 6.6.0 bis 6.9.0 kann es bei `SecretAgent`- und `RecordServer`-Servicekomponenten aufgrund einer falschen Konfiguration des Dateinamenmusters in den `Log4j2`-Eigenschaften zu einem Verlust von Protokolldaten kommen. Die falsche Konfiguration führt dazu, dass die Komponenten nur eine Protokolldatei pro Tag generieren. Wenn die Rotationsstrategie angewendet wird, überschreibt sie die vorhandene Datei, anstatt wie erwartet eine neue Protokolldatei zu generieren. Um das Problem zu umgehen, verwenden Sie eine Bootstrap-Aktion, um jede Stunde Protokolldateien zu generieren, und fügen Sie eine automatisch inkrementierte Ganzzahl an den Dateinamen an, um die Rotation zu handhaben.

Verwenden Sie für die Versionen Amazon EMR 6.6.0 bis 6.9.0 die folgende Bootstrap-Aktion, wenn Sie einen Cluster starten.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-6x/replace-puppet.sh,Args=[]"
```

Verwenden Sie für Amazon EMR 5.36.0 die folgende Bootstrap-Aktion, wenn Sie einen Cluster starten.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-5x/replace-puppet.sh,Args=[]"
```

## 6.6.0 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.4.1	Amazon SageMaker Spark-SDK
<code>emr-ddb</code>	4.16.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	3.2.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.5.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-notebook-env	1.5.0	Conda Env für EMR-Notebooks, das Jupyter Enterprise Gateway enthält
emr-s3-dist-cp	2.20.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	2.1.0	EMR S3Select-Konnektor
emrfs	2,50,0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1,14,2	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
flink-jobmanager-config	1.14.2	Verwaltung von Ressourcen auf EMR-Knoten für Apache JobManager Flink.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.



Komponente	Version	Beschreibung
hadoop-client	3.2.1-amzn-6	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	3.2.1-amzn-6	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	3.2.1-amzn-6	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	3.2.1-amzn-6	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-hdfs-journalnode	3.2.1-amzn-6	HDFS-Service zum Verwalten des Hadoop-Dateisystem journals auf HA-Clustern.
hadoop-httpfs-server	3.2.1-amzn-6	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	3.2.1-amzn-6	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	3.2.1-amzn-6	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	3.2.1-amzn-6	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.

Komponente	Version	Beschreibung
hadoop-yarn-resourcemanager	3.2.1-amzn-6	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	3.2.1-amzn-6	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	2.4.4-amzn-2	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	2.4.4-amzn-2	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	2.4.4-amzn-2	HBase-Befehlszeilen-Client.
hbase-rest-server	2.4.4-amzn-2	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	2.4.4-amzn-2	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hbase-operator-tools	2.4.4-amzn-2	Reparaturtool für Apache-HBase-Cluster.
hcatalog-client	3.1.2-amzn-7	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.

Komponente	Version	Beschreibung
hcatalog-server	3.1.2-amzn-7	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	3.1.2-amzn-7	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	3.1.2-amzn-7	Hive-Befehlszeilen-Client.
hive-hbase	3.1.2-amzn-7	Hive-hbase client.
hive-metastore-server	3.1.2-amzn-7	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	3.1.2-amzn-7	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.10.1-amzn-0	Inkrementelles Verarbeitungs-Framework zur Versorgung der Datenpipeline mit geringer Latenz und hoher Effizienz.
hudi-presto	0.10.1-amzn-0	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hudi-trino	0.10.1-amzn-0	Bündel-Bibliothek zum Ausführen von Trino mit Hudi.

Komponente	Version	Beschreibung
hudi-spark	0.10.1-amzn-0	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.
hue-server	4.10.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
iceberg	0.13.1	Apache Iceberg ist ein offenes Tabellenformat für sehr große analytische Datensätze
jupyterhub	1.4.1	Multi-User-Server für Jupyter-Notebooks
livy-server	0.7.1-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP- und Reverse-Proxy-Server.
mxnet	1.8.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.68+	MariaDB-Datenbankserver.
nvidia-cuda	11,0,194	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.2.1	Oozie-Befehlszeilen-Client.
oozie-server	5.2.1	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	4.5.0	Open Source Computer Vision Library.

Komponente	Version	Beschreibung
phoenix-library	5.1.2	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-connectors	5.1.2	Apache-Phoenix-Konnektoren für Spark-3
phoenix-query-server	5.1.2	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.267-amzn-0	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.267-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.267-amzn-0	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
trino-coordinator	367-amzn-0	Service zur Annahme von Abfragen und Verwaltung der Abfrageausführung der Trino-Worker.
trino-worker	367-amzn-0	Service für das Ausführen von Teilen einer Abfrage.

Komponente	Version	Beschreibung
trino-client	367-amzn-0	Trino-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Trino-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	4.0.2	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	2.0.0	Apache Ranger Key Management System
spark-client	3.2.0-amzn-0	Spark-Befehlszeilen-Clients.
spark-history-server	3.2.0-amzn-0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	3.2.0-amzn-0	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	3.2.0-amzn-0	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
spark-rapids	22.02.0-amzn-0	Nvidia-Spark-RAPIDS-Plugin, das Apache Spark mit GPUs beschleunigt.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.

Komponente	Version	Beschreibung
tensorflow	2.4.1	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.2	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.41+	Apache HTTP-Server.
zeppelin-server	0.10.0	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.5.7	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.5.7	ZooKeeper Befehlszeilen-Cli-ent.

## 6.6.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Aktionen zur Neukonfiguration treten auf, wenn Sie eine Konfiguration für Instance-Gruppen in einem laufenden Cluster angeben. Amazon EMR initiiert nur Rekonfigurationsaktionen für die Klassifizierungen, die Sie ändern. Weitere Informationen finden Sie unter [Eine Instance-Gruppe in einem laufenden Cluster neu konfigurieren](#).

## emr-6.6.0-Klassifizierungen

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
capacity-scheduler	Ändert die Werte in der capacity-scheduler.xml-Datei in Hadoop.	Restarts the ResourceManager service.
container-executor	Ändern Sie die Werte in der Datei „container-executor.cfg“ Datei von Hadoop YARN.	Not available.
container-log4j	Ändert die Werte in der container-log4j.properties-Datei in Hadoop YARN.	Not available.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Ändern Sie die Docker-bezogenen Einstellungen.	Not available.
emrfs-site	Ändert die EMRFS-Einstellungen.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the



Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
		Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegistrator, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.	Restarts Flink history server.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.	Restarts Flink history server.
flink-log4j-session	Ändern Sie die Einstellungen von Flink log4j-session.properties für Kubernetes-/Yarn-Sitzungen.	Restarts Flink history server.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.	Restarts Flink history server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.	Not available.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Ändert die Werte in der HBase-Umgebung.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.	Not available.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.	This classification should not be reconfigured.
hdfs-env	Ändert die Werte in der HDFS-Umgebung.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei proto-hive-site .xml von HCatalog.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.	Restarts Hive WebHCat server.
hive	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Hive.	Sets configurations to launch Hive LLAP service.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.	Not available.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.	Not available.
hive-env	Ändert die Werte in der Hive-Umgebung.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.	Not available.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log	Not available.
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.	Not available.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hue-ini	Ändert die Werte in der INI-Datei in Hue.	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.	Restarts Hadoop Httpfs service.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.	Not available.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.	Not available.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Ändern der Werte in der Hudi-Umgebung.	Not available.
hudi-defaults	Ändern Sie die Werte in der hudi-defaults.conf-Datei in Hudi.	Not available.
iceberg-defaults	Ändern Sie die Werte in der iceberg-defaults.conf-Datei von Iceberg.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
jupyter-notebook-conf	Ändert die Werte in der <code>jupyter_notebook_config.py</code> -Datei in Jupyter Notebook.	Not available.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei <code>jupyterhub_config.py</code> .	Not available.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.	Not available.
jupyter-sparkmagic-conf	Ändert die Werte in der <code>config.json</code> -Datei in Sparkmagic.	Not available.
livy-conf	Ändert die Werte in der <code>livy.conf</code> -Datei von Livy.	Restarts Livy Server.
livy-env	Ändert die Werte in der Livy-Umgebung.	Restarts Livy Server.
livy-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Livy.	Restarts Livy Server.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Ändern Sie die Werte in der Datei <code>mapred-site.xml</code> der MapReduce Anwendung.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Ändert die Werte in der Oozie-Umgebung.	Restarts Oozie.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.	Restarts Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.	Not available.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.	Not available.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.	Not available.
pig-env	Ändert die Werte in der Pig-Umgebung.	Not available.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.	Restarts Oozie.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.	Not available.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Ändert die Werte in der config.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)



Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei <code>password-authenticator.properties</code> .	Not available.
presto-env	Ändern Sie die Werte in der <code>presto-env.sh</code> -Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Ändern Sie die Werte in der <code>node.properties</code> -Datei in Presto.	Not available.
presto-connector-blackhole	Ändert die Werte in der <code>blackhole.properties</code> -Datei in Presto.	Not available.
presto-connector-cassandra	Ändert die Werte in der <code>cassandra.properties</code> -Datei in Presto.	Not available.
presto-connector-hive	Ändert die Werte in der <code>hive.properties</code> -Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Ändert die Werte in der <code>jmx.properties</code> -Datei in Presto.	Not available.
presto-connector-kafka	Ändert die Werte in der <code>kafka.properties</code> -Datei in Presto.	Not available.
presto-connector-localfile	Ändert die Werte in der <code>localfile.properties</code> -Datei in Presto.	Not available.
presto-connector-memory	Ändert die Werte in der <code>memory.properties</code> -Datei in Presto.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-mongodb	Ändert die Werte in der <code>mongodb.properties</code> -Datei in Presto.	Not available.
presto-connector-mysql	Ändert die Werte in der <code>mysql.properties</code> -Datei in Presto.	Not available.
presto-connector-postgresql	Ändert die Werte in der <code>postgresql.properties</code> -Datei in Presto.	Not available.
presto-connector-raptor	Ändert die Werte in der <code>raptor.properties</code> -Datei in Presto.	Not available.
presto-connector-redis	Ändert die Werte in der <code>redis.properties</code> -Datei in Presto.	Not available.
presto-connector-redshift	Ändert die Werte in der <code>redshift.properties</code> -Datei.	Not available.
presto-connector-tpch	Ändert die Werte in der <code>tpch.properties</code> -Datei in Presto.	Not available.
presto-connector-tpcds	Ändert die Werte in der <code>tpcds.properties</code> -Datei in Presto.	Not available.
trino-log	Ändern Sie die Werte in der <code>log.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)
trino-config	Ändern Sie die Werte in der <code>config.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-password-authenticator	Ändern Sie die Werte in der Trino-Datei <code>password-authenticator.properties</code> .	Restarts Trino-Server (for Trino)
trino-env	Ändern Sie die Werte in der <code>trino-env.sh</code> -Datei von Trino.	Restarts Trino-Server (for Trino)
trino-node	Ändern Sie die Werte in der <code>node.properties</code> -Datei in Trino.	Not available.
trino-connector-blackhole	Ändern Sie die Werte in der <code>blackhole.properties</code> -Datei in Trino.	Not available.
trino-connector-cassandra	Ändern Sie die Werte in der <code>cassandra.properties</code> -Datei in Trino.	Not available.
trino-connector-hive	Ändern Sie die Werte in der <code>hive.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Ändern Sie die Werte in der <code>iceberg.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Ändern Sie die Werte in der <code>jmx.properties</code> -Datei in Trino.	Not available.
trino-connector-kafka	Ändern Sie die Werte in der <code>kafka.properties</code> -Datei in Trino.	Not available.
trino-connector-localfile	Ändern Sie die Werte in der <code>localfile.properties</code> -Datei in Trino.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-connector-memory	Ändern Sie die Werte in der memory.properties-Datei in Trino.	Not available.
trino-connector-mongodb	Ändern Sie die Werte in der mongodb.properties-Datei in Trino.	Not available.
trino-connector-mysql	Ändern Sie die Werte in der mysql.properties-Datei in Trino.	Not available.
trino-connector-postgresql	Ändern Sie die Werte in der postgresql.properties-Datei in Trino.	Not available.
trino-connector-raptor	Ändern Sie die Werte in der raptor.properties-Datei in Trino.	Not available.
trino-connector-redis	Ändern Sie die Werte in der redis.properties-Datei in Trino.	Not available.
trino-connector-redshift	Ändern Sie die Werte in der redshift.properties-Datei in Trino.	Not available.
trino-connector-tpch	Ändern Sie die Werte in der tpch.properties-Datei in Trino.	Not available.
trino-connector-tpcds	Ändern Sie die Werte in der tpcds.properties-Datei in Trino.	Not available.
ranger-kms-dbks-site	Ändert die Werte in der dbks-site.xml-Datei von Ranger KMS.	Restarts Ranger KMS Server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Ändert die Werte in der Ranger KMS-Umgebung.	Restarts Ranger KMS Server.
ranger-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei von Ranger KMS.	Not available.
ranger-kms-db-ca	Ändert die Werte für die CA-Datei auf S3 für die MySQL SSL-Verbindung mit Ranger KMS.	Not available.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Ändert die Werte in der Spark-Umgebung.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.	Not available.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.	Not available.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.	Not available.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.	Not available.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.	Restart Oozie and HiveServer2.
yarn-env	Ändert die Werte in der YARN-Umgebung.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.	Restarts Zeppelin.
zeppelin-site	Ändern Sie die Konfigurationseinstellungen in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.	Restarts Zookeeper server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.	Restarts Zookeeper server.

## Amazon-EMR-Version 6.5.0

### 6.5.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

### Informationen zur Anwendungsversion

	emr-6.5.0	emr-6.4.0	emr-6.3.1	emr-6.3.0
AWS SDK for Java	1.12.31	1.12.31	1.11.977	1,11,977
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2.12,10	2.12.10	2.12.10	2.12.10

	emr-6.5.0	emr-6.4.0	emr-6.3.1	emr-6.3.0
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.14.0	1.13.1	1.12.1	1.12.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4.4	2.4.4	2.2.6	2.2.6
HCatalog	3.1.2	3.1.2	3.1.2	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.2	3.1.2	3.1.2	3.1.2
Hudi	0.9.0-amzn-1	0.8.0-amzn-0	0.7.0-amzn-0	0.7.0-amzn-0
Hue	4.9.0	4.9.0	4.9.0	4.9.0
Iceberg	0.12.0	-	-	-
JupyterEnterpriseGateway	2.1.0	2.1.0	2.1.0	2.1.0
JupyterHub	1.4.1	1.4.1	1.2.2	1.2.2
Livy	0.7.1	0.7.1	0.7.0	0.7.0
MXNet	1.8.0	1.8.0	1.7.0	1.7.0
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	5.1.2	5.1.2	5.0.0	5.0.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0



	emr-6.5.0	emr-6.4.0	emr-6.3.1	emr-6.3.0
Presto	0,261	0,254,1	0,245,1	0,245,1
Spark	3.1.2	3.1.2	3.1.1	3.1.1
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.4.1	2.4.1	2.4.1	2.4.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	360	359	350	350
Zeppelin	0.10.0	0.9.0	0.9.0	0.9.0
ZooKeeper	3,5.7	3.5.7	3.4.14	3.4.14

## 6.5.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 6.5.0. Änderungen beziehen sich auf Version 6.4.0.

Erste Version: 20. Januar 2022

Aktualisiertes Veröffentlichungsdatum: 21. März 2022

### Neue Features

- [Verwaltete Skalierung] Spark Shuffle Data Managed Scaling Optimization – Für Amazon-EMR-Versionen 5.34.0 und höher sowie EMR-Versionen 6.4.0 und höher unterstützt verwaltete Skalierung jetzt Spark Shuffle Data (Daten, die Spark partitionsübergreifend verteilt, um bestimmte Operationen auszuführen). Weitere Informationen zu Shuffle-Vorgängen finden Sie unter [Nutzung von Managed Scaling in Amazon EMR](#) im Verwaltungshandbuch für Amazon EMR und [Spark-Programmierungs-Handbuch](#).
- Ab Amazon EMR 5.32.0 und 6.5.0 ist die dynamische Executor-Größenanpassung für Apache Spark standardmäßig aktiviert. Sie können dieses Feature ein- oder ausschalten, indem Sie den Konfigurationsparameter `spark.yarn.heterogeneousExecutors.enabled` verwenden.
- Support für das offene Tabellenformat von Apache Iceberg für riesige analytische Datensätze.

- Support für ranger-trino-plugin 2.0.1-amzn-1
- Support für toree 0.5.0

## Änderungen, Verbesserungen und behobene Probleme

- Die Release-Version von Amazon EMR 6.5 unterstützt jetzt Apache Iceberg 0.12.0 und bietet Laufzeitverbesserungen mit Amazon-EMR-Laufzeit für Apache Spark, Amazon-EMR-Laufzeit für Presto und Amazon-EMR-Laufzeit für Apache Hive.
- [Apache Iceberg](#) ist ein offenes Tabellenformat für große Datensätze in Amazon S3 und bietet schnelle Abfrageleistung bei großen Tabellen, atomare Commits, gleichzeitige Schreibvorgänge und eine SQL-kompatible Tabellenentwicklung. Mit EMR 6.5 können Sie Apache Spark 3.1.2 mit dem Iceberg-Tabellenformat verwenden.
- Apache Hudi 0.9 fügt Spark SQL DDL- und DML-Unterstützung hinzu. Auf diese Weise können Sie Hudi-Tabellen nur mit SQL-Anweisungen erstellen und ändern. Apache Hudi 0.9 beinhaltet auch Leistungsverbesserungen auf der Abfrageseite und auf der Writer-Seite.
- Amazon-EMR-Laufzeit für Apache Hive verbessert die Leistung von Apache Hive auf Amazon S3, indem Umbenennungsvorgänge bei Staging-Vorgängen entfernt werden, und verbessert die Leistung von Metastore Check (MSCK)-Befehlen, die zur Reparatur von Tabellen verwendet werden.

## Bekannte Probleme

- Wenn Amazon-EMR-Version 6.5.0, 6.6.0 oder 6.7.0 Apache Phoenix-Tabellen über die Apache Spark-Shell liest, tritt `NoSuchMethodError` ein, weil Amazon EMR ein falsches `Hbase.compat.version` verwendet. Amazon EMR Version 6.8.0 behebt dieses Problem.
- Hbase-Bundle-Cluster mit hoher Verfügbarkeit (HA) können nicht mit der Standard-Volume-Größe und dem Instance-Typ bereitgestellt werden. Dieses Problem lässt sich umgehen, indem Sie die Größe des Root-Volumens erhöhen.
- Um Spark-Aktionen mit Apache Oozie zu verwenden, müssen Sie Ihrer Oozie-`workflow.xml`-Datei die folgende Konfiguration hinzufügen. Andernfalls fehlen mehrere wichtige Bibliotheken wie Hadoop und EMRFS im Klassenpfad der Spark-Executoren, die Oozie startet.

```
<spark-opts>--conf spark.yarn.populateHadoopClasspath=true</spark-opts>
```

- Wenn Sie Spark mit der Formatierung des Hive-Partitionsstandorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den Amazon-EMR-Versionen 5.30.0 bis 5.36.0 und 6.2.0

bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:

- Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.
- Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist `s3://bucket/table/p=a` ein Präfix von `s3://bucket/table/p=a b`.
- Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen UTF-8-Wert, der kleiner als das `/`-Zeichen (U+002F) ist. Beispielsweise fällt das Leerzeichen (U+0020), das in `s3://bucket/table/p=a b` zwischen `a` und `b` vorkommt, in diese Kategorie. Beachten Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: `!"#$%&'()*+,-.` Weitere Informationen finden Sie unter [UTF-8-Kodierungstabelle und Unicode-Zeichen](#).

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.

## 6.5.0 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.4.1	Amazon SageMaker Spark-SDK

Komponente	Version	Beschreibung
emr-ddb	4.16.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	3.2.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.5.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-notebook-env	1.4.0	Conda Env für EMR-Notebooks, das Jupyter Enterprise Gateway enthält
emr-s3-dist-cp	2.19.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	2.1.0	EMR S3Select-Konnektor
emrfs	2,48,0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.14.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
flink-jobmanager-config	1.14.0	Verwaltung von Ressourcen auf EMR-Knoten für Apache JobManager Flink.

Komponente	Version	Beschreibung
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	3.2.1-amzn-5	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	3.2.1-amzn-5	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	3.2.1-amzn-5	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	3.2.1-amzn-5	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-hdfs-journalnode	3.2.1-amzn-5	HDFS-Service zum Verwalten des Hadoop-Dateisystem journals auf HA-Clustern.

Komponente	Version	Beschreibung
hadoop-httfs-server	3.2.1-amzn-5	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	3.2.1-amzn-5	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	3.2.1-amzn-5	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	3.2.1-amzn-5	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	3.2.1-amzn-5	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	3.2.1-amzn-5	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	2.4.4-amzn-1	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	2.4.4-amzn-1	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.

Komponente	Version	Beschreibung
hbase-client	2.4.4-amzn-1	HBase-Befehlszeilen-Client.
hbase-rest-server	2.4.4-amzn-1	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	2.4.4-amzn-1	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	3.1.2-amzn-6	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	3.1.2-amzn-6	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	3.1.2-amzn-6	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	3.1.2-amzn-6	Hive-Befehlszeilen-Client.
hive-hbase	3.1.2-amzn-6	Hive-hbase client.
hive-metastore-server	3.1.2-amzn-6	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	3.1.2-amzn-6	Service zur Annahme von Hive-Abfragen als Webanfragen.

Komponente	Version	Beschreibung
hudi	0.9.0-amzn-1	Inkrementelles Verarbeitungs-Framework zur Versorgung der Datenpipeline mit geringer Latenz und hoher Effizienz.
hudi-presto	0.9.0-amzn-1	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hudi-trino	0.9.0-amzn-1	Bündel-Bibliothek zum Ausführen von Trino mit Hudi.
hudi-spark	0.9.0-amzn-1	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.
hue-server	4.9.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
iceberg	0.12.0	Apache Iceberg ist ein offenes Tabellenformat für sehr große analytische Datensätze
jupyterhub	1.4.1	Multi-User-Server für Jupyter-Notebooks
livy-server	0.7.1-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP- und Reverse-Proxy-Server.
mxnet	1.8.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.68+	MariaDB-Datenbankserver.



Komponente	Version	Beschreibung
nvidia-cuda	10.1,243	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.2.1	Oozie-Befehlszeilen-Client.
oozie-server	5.2.1	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	5.1.2	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	5.1.2	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.261-amzn-0	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.261-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.261-amzn-0	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.

Komponente	Version	Beschreibung
trino-coordinator	360	Service zur Annahme von Abfragen und Verwaltung der Abfrageausführung der Trino-Worker.
trino-worker	360	Service für das Ausführen von Teilen einer Abfrage.
trino-client	360	Trino-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Trino-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	4.0.2	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	2.0.0	Apache Ranger Key Management System
spark-client	3.1.2-amzn-1	Spark-Befehlszeilen-Clients.
spark-history-server	3.1.2-amzn-1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	3.1.2-amzn-1	In-Memory-Ausführungs-Engine für YARN.

Komponente	Version	Beschreibung
spark-yarn-slave	3.1.2-amzn-1	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
spark-rapids	0.4.1	Nvidia-Spark-RAPIDS-Plugin, das Apache Spark mit GPUs beschleunigt.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	2.4.1	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.2	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.41+	Apache HTTP-Server.
zeppelin-server	0.10.0	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.5.7	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.5.7	ZooKeeper Befehlszeilen-Client.

## 6.5.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Aktionen zur Neukonfiguration treten auf, wenn Sie eine Konfiguration für Instance-Gruppen in einem laufenden Cluster angeben. Amazon EMR initiiert nur Rekonfigurationsaktionen für die Klassifizierungen, die Sie ändern. Weitere Informationen finden Sie unter [Eine Instance-Gruppe in einem laufenden Cluster neu konfigurieren](#).

### emr-6.5.0-Klassifizierungen

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.	Restarts the Resource Manager service.
container-executor	Ändern Sie die Werte in der Datei „ <code>container-executor.cfg</code> “ Datei von Hadoop YARN.	Not available.
container-log4j	Ändert die Werte in der <code>container-log4j.properties</code> -Datei in Hadoop YARN.	Not available.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
		Httpfs, and MapReduce-HistoryServer.
docker-conf	Ändern Sie die Docker-bezogenen Einstellungen.	Not available.
emrfs-site	Ändert die EMRFS-Einstellungen.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegistratorserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.	Restarts Flink history server.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.	Restarts Flink history server.
flink-log4j-session	Ändern Sie die Einstellungen von Flink log4j-session.properties für Kubernetes-/Yarn-Sitzungen.	Restarts Flink history server.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.	Restarts Flink history server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.	Not available.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Ändert die Werte in der HBase-Umgebung.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.	Not available.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.	This classification should not be reconfigured.
hdfs-env	Ändert die Werte in der HDFS-Umgebung.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei proto-hive-site .xml von HCatalog.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.	Restarts Hive WebHCat server.
hive	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Hive.	Sets configurations to launch Hive LLAP service.



Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.	Not available.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.	Not available.
hive-env	Ändert die Werte in der Hive-Umgebung.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.	Not available.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log	Not available.
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.	Not available.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hue-ini	Ändert die Werte in der INI-Datei in Hue.	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.	Restarts Hadoop Httpfs service.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.	Not available.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.	Not available.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Ändern der Werte in der Hudi-Umgebung.	Not available.
hudi-defaults	Ändern Sie die Werte in der hudi-defaults.conf-Datei in Hudi.	Not available.
iceberg-defaults	Ändern Sie die Werte in der iceberg-defaults.conf-Datei von Iceberg.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
jupyter-notebook-conf	Ändert die Werte in der <code>jupyter_notebook_config.py</code> -Datei in Jupyter Notebook.	Not available.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei <code>jupyterhub_config.py</code> .	Not available.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.	Not available.
jupyter-sparkmagic-conf	Ändert die Werte in der <code>config.json</code> -Datei in Sparkmagic.	Not available.
livy-conf	Ändert die Werte in der <code>livy.conf</code> -Datei von Livy.	Restarts Livy Server.
livy-env	Ändert die Werte in der Livy-Umgebung.	Restarts Livy Server.
livy-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Livy.	Restarts Livy Server.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Ändern Sie die Werte in der Datei <code>mapred-site.xml</code> der MapReduce Anwendung.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Ändert die Werte in der Oozie-Umgebung.	Restarts Oozie.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.	Restarts Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.	Not available.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.	Not available.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.	Not available.
pig-env	Ändert die Werte in der Pig-Umgebung.	Not available.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.	Restarts Oozie.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.	Not available.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Ändert die Werte in der config.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei <code>password-authenticator.properties</code> .	Not available.
presto-env	Ändern Sie die Werte in der <code>presto-env.sh</code> -Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Ändern Sie die Werte in der <code>node.properties</code> -Datei in Presto.	Not available.
presto-connector-blackhole	Ändert die Werte in der <code>blackhole.properties</code> -Datei in Presto.	Not available.
presto-connector-cassandra	Ändert die Werte in der <code>cassandra.properties</code> -Datei in Presto.	Not available.
presto-connector-hive	Ändert die Werte in der <code>hive.properties</code> -Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Ändert die Werte in der <code>jmx.properties</code> -Datei in Presto.	Not available.
presto-connector-kafka	Ändert die Werte in der <code>kafka.properties</code> -Datei in Presto.	Not available.
presto-connector-localfile	Ändert die Werte in der <code>localfile.properties</code> -Datei in Presto.	Not available.
presto-connector-memory	Ändert die Werte in der <code>memory.properties</code> -Datei in Presto.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-mongodb	Ändert die Werte in der <code>mongodb.properties</code> -Datei in Presto.	Not available.
presto-connector-mysql	Ändert die Werte in der <code>mysql.properties</code> -Datei in Presto.	Not available.
presto-connector-postgresql	Ändert die Werte in der <code>postgresql.properties</code> -Datei in Presto.	Not available.
presto-connector-raptor	Ändert die Werte in der <code>raptor.properties</code> -Datei in Presto.	Not available.
presto-connector-redis	Ändert die Werte in der <code>redis.properties</code> -Datei in Presto.	Not available.
presto-connector-redshift	Ändert die Werte in der <code>redshift.properties</code> -Datei.	Not available.
presto-connector-tpch	Ändert die Werte in der <code>tpch.properties</code> -Datei in Presto.	Not available.
presto-connector-tpcds	Ändert die Werte in der <code>tpcds.properties</code> -Datei in Presto.	Not available.
trino-log	Ändern Sie die Werte in der <code>log.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)
trino-config	Ändern Sie die Werte in der <code>config.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-password-authenticator	Ändern Sie die Werte in der Trino-Datei <code>password-authenticator.properties</code> .	Restarts Trino-Server (for Trino)
trino-env	Ändern Sie die Werte in der <code>trino-env.sh</code> -Datei von Trino.	Restarts Trino-Server (for Trino)
trino-node	Ändern Sie die Werte in der <code>node.properties</code> -Datei in Trino.	Not available.
trino-connector-blackhole	Ändern Sie die Werte in der <code>blackhole.properties</code> -Datei in Trino.	Not available.
trino-connector-cassandra	Ändern Sie die Werte in der <code>cassandra.properties</code> -Datei in Trino.	Not available.
trino-connector-hive	Ändern Sie die Werte in der <code>hive.properties</code> -Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Ändern Sie die Werte in der <code>jmx.properties</code> -Datei in Trino.	Not available.
trino-connector-kafka	Ändern Sie die Werte in der <code>kafka.properties</code> -Datei in Trino.	Not available.
trino-connector-localfile	Ändern Sie die Werte in der <code>localfile.properties</code> -Datei in Trino.	Not available.
trino-connector-memory	Ändern Sie die Werte in der <code>memory.properties</code> -Datei in Trino.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-connector-mongodb	Ändern Sie die Werte in der <code>mongodb.properties</code> -Datei in Trino.	Not available.
trino-connector-mysql	Ändern Sie die Werte in der <code>mysql.properties</code> -Datei in Trino.	Not available.
trino-connector-postgresql	Ändern Sie die Werte in der <code>postgresql.properties</code> -Datei in Trino.	Not available.
trino-connector-raptor	Ändern Sie die Werte in der <code>raptor.properties</code> -Datei in Trino.	Not available.
trino-connector-redis	Ändern Sie die Werte in der <code>redis.properties</code> -Datei in Trino.	Not available.
trino-connector-redshift	Ändern Sie die Werte in der <code>redshift.properties</code> -Datei in Trino.	Not available.
trino-connector-tpch	Ändern Sie die Werte in der <code>tpch.properties</code> -Datei in Trino.	Not available.
trino-connector-tpcds	Ändern Sie die Werte in der <code>tpcds.properties</code> -Datei in Trino.	Not available.
ranger-kms-dbks-site	Ändert die Werte in der <code>dbks-site.xml</code> -Datei von Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Ändern Sie die Werte in der <code>ranger-kms-site</code> XML-Datei von Ranger KMS.	Restarts Ranger KMS Server.



Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
ranger-kms-env	Ändert die Werte in der Ranger KMS-Umgebung.	Restarts Ranger KMS Server.
ranger-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei von Ranger KMS.	Not available.
ranger-kms-db-ca	Ändert die Werte für die CA-Datei auf S3 für die MySQL SSL-Verbindung mit Ranger KMS.	Not available.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Ändert die Werte in der Spark-Umgebung.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.	Not available.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.	Not available.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.	Not available.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.	Restart Oozie and HiveServer2.
yarn-env	Ändert die Werte in der YARN-Umgebung.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.	Restarts Zeppelin.
zeppelin-site	Ändern Sie die Konfigurationseinstellungen in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.	Restarts Zookeeper server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.	Restarts Zookeeper server.

## Amazon-EMR-Version 6.4.0

### 6.4.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

### Informationen zur Anwendungsversion

	emr-6.4.0	emr-6.3.1	emr-6.3.0	emr-6.2.1
AWS SDK for Java	1.12.31	1.11.977	1,11,977	1.11.880
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2.12,10	2.12.10	2.12.10	2.12.10

	emr-6.4.0	emr-6.3.1	emr-6.3.0	emr-6.2.1
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.13.1	1.12.1	1.12.1	1.11.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4.4	2.2.6	2.2.6	2.2.6-amzn-0
HCatalog	3.1.2	3.1.2	3.1.2	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.2	3.1.2	3.1.2	3.1.2
Hudi	0.8.0-amzn-0	0.7.0-amzn-0	0.7.0-amzn-0	0.6.0-amzn-1
Hue	4.9.0	4.9.0	4.9.0	4.8.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	2.1.0	2.1.0	2.1.0	2.1.0
JupyterHub	1.4.1	1.2.2	1.2.2	1.1.0
Livy	0.7.1	0.7.0	0.7.0	0.7.0
MXNet	1.8.0	1.7.0	1.7.0	1.7.0
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.0
Phoenix	5.1.2	5.0.0	5.0.0	5.0.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0

	emr-6.4.0	emr-6.3.1	emr-6.3.0	emr-6.2.1
Presto	0,254,1	0,245,1	0,245,1	0,238,3
Spark	3.1.2	3.1.1	3.1.1	3.0.1
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.4.1	2.4.1	2.4.1	2.3.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	359	350	350	343
Zeppelin	0.9.0	0.9.0	0.9.0	0.9.0
ZooKeeper	3,5.7	3.4.14	3.4.14	3.4.14

## 6.4.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 6.4.0. Änderungen beziehen sich auf Version 6.3.0.

Erste Version: 20. September 2021

Aktualisiertes Veröffentlichungsdatum: 21. März 2022

### Unterstützte Anwendungen

- AWS SDK for Java Ausführung 1.12.31
- CloudWatch Version 2.2.0 senken
- Version 4.16.0 von EMR DynamoDB Connector
- EMRFS-Version 2.47.0
- Version 3.2.0 von Amazon EMR Goodies
- Version 3.5.0 von Amazon EMR Kinesis Connector
- Amazon-EMR-Aufzeichnungsserver Version 2.1.0
- Version 2.5.0 von Amazon EMR Skripts

- Flink, Version 1.13.1
- Ganglia Version 3.7.2
- AWS Glue Hive Metastore Client Version 3.3.0
- Hadoop Version 3.2.1-amzn-4
- HBase Version 2.4.4-amzn-0
- ase-operator-tools HB 1.1.0
- HCatalog Version 3.1.2-amzn-5
- Hive Version 3.1.2-amzn-5
- Hudi Version 0.8.0-amzn-0
- Hue Version 4.9.0
- Java JDK-Version Corretto-8.302.08.1 (Build 1.8.0\_302-b08)
- JupyterHub Ausführung 1.4.1
- Livy Version 0.7.1-incubating
- MXNet-Version 1.8.0
- Oozie-Version 5.2.1
- Phoenix Version 5.1.2
- Pig Version 0.17.0
- Presto Version 0.254.1-amzn-0
- Trino Version 359
- Version 2.0.0 von Apache Ranger KMS (transparente Multi-Master-Verschlüsselung)
- Ranger-Plugins 2.0.1-amzn-0
- Ranger-S3-Plugin 1.2.0
- SageMaker Spark SDK, Version 1.4.1
- Scala-Version 2.12.10 (OpenJDK 64-Bit-Server-VM, Java 1.8.0\_282)
- Spark Version 3.1.2-amzn-0
- Spark-Rapids 0.4.1
- Sqoop-Version 1.4.7
- TensorFlow Ausführung 2.4.1
- Tez-Version 0.9.2
- Zeppelin-Version 0.9.0

- Zookeeper Version 3.5.7
- Konnektoren und Treiber: DynamoDB Connector 4.16.0

## Neue Features

- [Verwaltete Skalierung] Spark Shuffle Data Managed Scaling Optimization – Für Amazon-EMR-Versionen 5.34.0 und höher sowie EMR-Versionen 6.4.0 und höher unterstützt verwaltete Skalierung jetzt Spark Shuffle Data (Daten, die Spark partitionsübergreifend verteilt, um bestimmte Operationen auszuführen). Weitere Informationen zu Shuffle-Vorgängen finden Sie unter [Nutzung von Managed Scaling in Amazon EMR](#) im Verwaltungshandbuch für Amazon EMR und [Spark-Programmierungs-Handbuch](#).
- Auf Apache-Ranger-fähigen Amazon-EMR-Clustern können Sie Apache Spark SQL verwenden, um Daten mit, INSERT INTO, INSERT OVERWRITE und ALTER TABLE in die Apache-Hive-Metastore-Tabellen einzufügen oder diese zu aktualisieren. Wenn Sie ALTER TABLE mit Spark SQL verwenden, muss ein Partitionsspeicherort das untergeordnete Verzeichnis eines Tabellenspeicherorts sein. Amazon EMR unterstützt derzeit nicht das Einfügen von Daten in eine Partition, deren Partitionsspeicherort sich von der Tabellenposition unterscheidet.
- PrestoSQL wurde [in Trino umbenannt](#).
- Hive: Die Ausführung einfacher SELECT-Abfragen mit LIMIT-Klausel wird beschleunigt, indem die Abfrageausführung gestoppt wird, sobald die in der LIMIT-Klausel angegebene Anzahl von Datensätzen abgerufen wurde. Einfache SELECT-Abfragen sind Abfragen ohne GROUP BY/ORDER BY-Klausel oder Abfragen ohne Reducer-Stage. z. B. SELECT \* from <TABLE> WHERE <Condition> LIMIT <Number>.

## Hudi-Parallelitätskontrolle

- Hudi unterstützt jetzt Optimistic Concurrency Control (OCC – Optimistische Parallelitätskontrolle), die mit Schreiboperationen wie UPSERT und INSERT genutzt werden kann, um Änderungen von mehreren Autoren an derselben Hudi-Tabelle zu ermöglichen. Dies ist OCC auf Dateiebene, sodass zwei beliebige Commits (oder Writer) in dieselbe Tabelle schreiben können, sofern ihre Änderungen nicht kollidieren. Weitere Informationen finden Sie unter [Hudi-Parallelitätskontrolle..](#)
- Auf Amazon-EMR-Clustern ist Zookeeper installiert, der als Sperranbieter für OCC genutzt werden kann. Um die Verwendung dieses Features zu vereinfachen, sind in Amazon-EMR-Clustern die folgenden Eigenschaften vorkonfiguriert:

```
hoodie.write.lock.provider=org.apache.hudi.client.transaction.lock.ZookeeperBasedLockProvider
```

```
hoodie.write.lock.zookeeper.url=<EMR Zookeeper URL>
hoodie.write.lock.zookeeper.port=<EMR Zookeeper Port>
hoodie.write.lock.zookeeper.base_path=/hudi
```

Um OCC zu aktivieren, müssen Sie die folgenden Eigenschaften entweder mit ihren Hudi-Auftragsoptionen oder auf Cluster-Ebene mithilfe der Amazon-EMR-Konfigurations-API konfigurieren:

```
hoodie.write.concurrency.mode=optimistic_concurrency_control
hoodie.cleaner.policy.failed.writes=LAZY (Performs cleaning of failed writes lazily
instead of inline with every write)
hoodie.write.lock.zookeeper.lock_key=<Key to uniquely identify the Hudi table> (Table
Name is a good option)
```

### Hudi Monitoring: CloudWatch Amazon-Integration zur Berichterstattung über Hudi-Metriken

- Amazon EMR unterstützt die Veröffentlichung von Hudi Metrics auf Amazon. CloudWatch Es wird aktiviert, indem die folgenden erforderlichen Konfigurationen festgelegt werden:

```
hoodie.metrics.on=true
hoodie.metrics.reporter.type=CLOUDWATCH
```

- Im Folgenden finden Sie optionale Hudi-Konfigurationen, die Sie ändern können:

Einstellung	Beschreibung	Wert
hoodie.metrics.cloudwatch.report.period.seconds	Häufigkeit (in Sekunden), mit der Kennzahlen an Amazon gemeldet werden CloudWatch	Der Standardwert ist 60s, was für die von Amazon angebotene Standardauflösung von einer Minute in Ordnung ist CloudWatch
hoodie.metrics.cloudwatch.metric.prefix	Präfix, das jedem Metriknamen hinzugefügt werden soll	Der Standardwert ist leer (kein Präfix)
hoodie.metrics.cloudwatch.namespace	CloudWatch Amazon-Namespace, unter dem Metriken veröffentlicht werden	Der Standardwert ist Hudi



Einstellung	Beschreibung	Wert
<code>hoodie.metrics.cloudwatch.maxDatumsPerAnfrage</code>	Maximale Anzahl von Daten, die in einer Anfrage an Amazon enthalten sein können CloudWatch	Der Standardwert ist 20, was dem CloudWatch Amazon-Standard entspricht

## Unterstützung und Verbesserungen für Amazon-EMR-Hudi-Konfigurationen

- Kunden können jetzt die EMR-Konfigurations-API und das Rekonfigurationsfeature nutzen, um Hudi-Konfigurationen auf Cluster-Ebene zu konfigurieren. Eine neue dateibasierte Konfigurationsunterstützung wurde über `/etc/hudi/conf/hudi-defaults.conf` eingeführt, ähnlich wie bei anderen Anwendungen wie Spark, Hive usw. EMR konfiguriert einige Standardwerte, um die Benutzererfahrung zu verbessern:
  - `hoodie.datasource.hive_sync.jdbcurl` ist für die Cluster-Hive-Server-URL konfiguriert und muss nicht mehr angegeben werden. Dies ist besonders nützlich, wenn Sie einen Auftrag im Spark-Cluster-Modus ausführen, wo Sie zuvor die Amazon-EMR-Master-IP angeben mussten.
  - HBase-spezifische Konfigurationen, die für die Verwendung des HBase-Index mit Hudi nützlich sind.
  - Spezifische Konfiguration des Zookeeper-Sperranbieters, wie unter Parallelitätskontrolle beschrieben, was die Verwendung von Optimistischer Parallelitätskontrolle (OCC) erleichtert.
- Zusätzliche Änderungen wurden eingeführt, um die Anzahl der Konfigurationen zu reduzieren, die Sie bestehen müssen, und um nach Möglichkeit automatische Schlüsse zu ziehen:
  - Das `partitionBy` -Schlüsselwort kann verwendet werden, um die Partitionsspalte zu spezifizieren.
  - Bei der Aktivierung von Hive Sync ist es nicht mehr erforderlich, den Vorgang `HIVE_TABLE_OPT_KEY`, `HIVE_PARTITION_FIELDS_OPT_KEY`, `HIVE_PARTITION_EXTRACTOR_CLASS_OPT_KEY` zu bestehen. Diese Werte können aus dem Hudi-Tabellennamen und dem Partitionsfeld abgeleitet werden.
  - `KEYGENERATOR_CLASS_OPT_KEY` ist nicht zwingend erforderlich und kann aus einfacheren Fällen von `SimpleKeyGenerator` und `ComplexKeyGenerator` abgeleitet werden.

## Vorbehalte von Hudi

- Hudi unterstützt keine vektorisierte Ausführung in Hive für Merge on Read (MoR)- und Bootstrap-Tabellen. So schlägt beispielsweise `count( *)` mit der Hudi-Echtzeittabelle fehl, wenn `hive.vectorized.execution.enabled` auf „wahr“ gesetzt ist. Um das Problem zu umgehen, können Sie das vektorisierte Lesen deaktivieren, indem Sie `hive.vectorized.execution.enabled` auf `false` setzen.
- Die Multi-Writer-Unterstützung ist nicht mit dem Hudi-Bootstrap-Feature kompatibel.
- Flink Streamer und Flink SQL sind experimentelle Features in dieser Version. Diese Features werden nicht zur Verwendung in Produktionsbereitstellungen empfohlen.

## Änderungen, Verbesserungen und behobene Probleme

Dies ist eine Version zur Behebung von Problemen mit Amazon-EMR-Skalierung, wenn ein Cluster nicht erfolgreich hoch-/herunterskaliert werden kann oder Anwendungsfehler verursacht werden.

- Zuvor führte ein manueller Neustart des Ressourcenmanagers auf einem Multi-Master-Cluster dazu, dass Amazon-EMR-Cluster-Daemons wie Zookeeper alle zuvor stillgelegten oder verlorenen Knoten in der Zookeeper `znode`-Datei neu geladen haben. Dies führte dazu, dass die Standardgrenzwerte in bestimmten Situationen überschritten wurden. Amazon EMR entfernt jetzt die außer Betrieb genommenen oder verlorenen Knotendatensätze, die älter als eine Stunde sind, aus der Zookeeper-Datei, und die internen Grenzwerte wurden erhöht.
- Es wurde ein Problem behoben, bei dem Skalierungsanforderungen für einen großen, stark ausgelasteten Cluster fehlschlagen, wenn Amazon-EMR-On-Cluster-Daemons Aktivitäten zur Integritätsprüfung durchführten, z. B. das Erfassen des YARN-Knotenstatus und des HDFS-Knotenstatus. Das lag daran, dass Cluster-Daemons die Integritätsstatusdaten eines Knotens nicht an interne Amazon-EMR-Komponenten weitergeben konnten.
- Verbesserte EMR-On-Cluster-Daemons zur korrekten Nachverfolgung der Knotenstatus bei der Wiederverwendung von IP-Adressen, um die Zuverlässigkeit bei Skalierungsvorgängen zu verbessern.
- [SPARK-29683](#). Es wurde ein Problem behoben, bei dem während der Cluster-Skalierung Auftragsfehler auftraten, da Spark davon ausging, dass alle verfügbaren Knoten auf der Verweigern-Liste standen.
- [YARN-9011](#). Es wurde ein Problem behoben, bei dem Auftragsfehler aufgrund eines Fehlers bei der Außerbetriebnahme von YARN auftraten, wenn der Cluster versuchte, hoch- oder herunterzuskalieren.

- Das Problem mit Schritt- oder Auftragsfehlern bei der Cluster-Skalierung wurde behoben, indem sichergestellt wurde, dass die Knotenstatus zwischen den Amazon-EMR-On-Cluster-Daemons und YARN/HDFS immer konsistent sind.
- Es wurde ein Problem behoben, bei dem Clustervorgänge wie Herunterskalierung und Schrittübermittlung für Amazon-EMR-Cluster, die mit Kerberos-Authentifizierung aktiviert waren, fehlschlagen. Dies lag daran, dass der Amazon-EMR-On-Cluster-Daemon das Kerberos-Ticket nicht erneuert hat, das für die sichere Kommunikation mit HDFS/YARN erforderlich ist, das auf dem Primärknoten ausgeführt wird.
- Konfigurieren eines Clusters zur Behebung von Leistungsproblemen mit Apache YARN Timeline Server Version 1 und 1.5

Apache YARN Timeline Server Version 1 und 1.5 können bei sehr aktiven, großen EMR-Clustern zu Leistungsproblemen führen, insbesondere bei `yarn.resourcemanager.system-metrics-publisher.enabled=true`, was die Standardeinstellung in Amazon EMR ist. Ein Open-Source-YARN-Timeline-Server v2 löst das Leistungsproblem im Zusammenhang mit der Skalierbarkeit des YARN Timeline Servers.

Weitere Lösungen für dieses Problem umfassen:

- Konfiguration von `yarn.resourcemanager.system-metrics-publisher.enabled=false` in `yarn-site.xml`.
- Aktivieren des Fixes für dieses Problem beim Erstellen eines Clusters, wie unten beschrieben.

Die folgenden Amazon-EMR-Versionen enthalten eine Lösung für dieses Leistungsproblem mit YARN Timeline Server.

EMR 5.30.2, 5.31.1, 5.32.1, 5.33.1, 5.34.x, 6.0.1, 6.1.1, 6.2.1, 6.3.1, 6.4.x

Um den Fix für eine der oben angegebenen Amazon-EMR-Versionen zu aktivieren, legen Sie diese Eigenschaften `true` in einer JSON-Konfigurationsdatei fest, die mit dem [aws emr create-cluster-Befehlsparameter](#) `--configurations file://./configurations.json` übergeben wird. Oder aktivieren Sie den Fix über die Benutzeroberfläche der [Rekonfigurationskonsole](#).

Beispiel für den Inhalt der Datei `configurations.json`:

```
[
{
  "Classification": "yarn-site",
```

```
"Properties": {
  "yarn.resourcemanager.system-metrics-publisher.timeline-server-v1.enable-batch":
    "true",
  "yarn.resourcemanager.system-metrics-publisher.enabled": "true"
},
"Configurations": []
}
```

- WebHDFS- und HttpFS-Server werden standardmäßig deaktiviert. Sie können WebHDFS mithilfe der Hadoop-Konfiguration `dfs.webhdfs.enabled` erneut aktivieren. Der HttpFS-Server kann mit `sudo systemctl start hadoop-httpfs` gestartet werden.
- HTTPS ist jetzt standardmäßig für Amazon-Linux-Repositorys aktiviert. Wenn Sie eine Amazon-S3-VPCE-Richtlinie verwenden, um den Zugriff auf bestimmte Buckets zu beschränken, müssen Sie den neuen Amazon-Linux-Bucket-ARN `arn:aws:s3:::amazonlinux-2-repos-$region/*` zu Ihrer Richtlinie hinzufügen (ersetzen Sie `$region` durch die Region, in der sich der Endpunkt befindet). Weitere Informationen finden Sie unter diesem Thema in den Diskussionsforen. [AWS Ankündigung: Amazon Linux 2 unterstützt jetzt die Möglichkeit, HTTPS zu verwenden, wenn eine Verbindung zu Paket-Repositorys hergestellt wird.](#)
- Hive: Die Leistung bei Schreibabfragen wurde verbessert, indem die Verwendung eines Scratch-Verzeichnisses auf HDFS für den letzten JAuftrag aktiviert wurde. Die temporären Daten für den endgültigen Auftrag werden in HDFS statt in Amazon S3 geschrieben, und die Leistung wurde verbessert, da die Daten von HDFS an den endgültigen Tabellenort (Amazon S3) und nicht zwischen Amazon-S3-Geräten verschoben werden.
- Hive: Verbesserung der Kompilierungszeit von Abfragen um das 2,5-fache mit Glue Metastore Partition Pruning.
- Wenn integrierte UDFs von Hive an den Hive-Metastore-Server übergeben werden, wird standardmäßig nur eine Teilmenge dieser integrierten UDFs an den Glue Metastore übergeben, da Glue nur begrenzte Ausdrucksoperatoren unterstützt. Wenn Sie `hive.glue.partition.pruning.client=true` festlegen, erfolgt das gesamte Partitionsbereinigen auf der Clientseite. Wenn Sie `hive.glue.partition.pruning.server=true` festlegen, erfolgt das gesamte Bereinigen von Partitionen auf der Serverseite.

## Bekannte Probleme

- Hue-Abfragen funktionieren in Amazon EMR 6.4.0 nicht, da der Apache-Hadoop-HttpFS-Server standardmäßig deaktiviert ist. Um Hue auf Amazon EMR 6.4.0 zu verwenden, starten Sie den HttpFS-Server auf dem Amazon-EMR-Primärknoten entweder manuell mithilfe von `sudo systemctl start hadoop-httpfs` oder [verwenden Sie einen Amazon-EMR-Schritt](#).
- Das Amazon-EMR-Notebooks-Feature, das mit dem Livy-Benutzerwechsel verwendet wird, funktioniert nicht, da HttpFS standardmäßig deaktiviert ist. In diesem Fall kann das EMR Notebook keine Verbindung zu dem Cluster herstellen, für den Livy-Identitätswechsel aktiviert ist. Die Problemlösung besteht darin, den HttpFS-Server zu starten, bevor Sie das EMR-Notebook mit dem Cluster `sudo systemctl start hadoop-httpfs` verbinden.
- In Amazon-EMR-Version 6.4.0 unterstützt Phoenix die Komponente Phoenix-Konnektoren nicht.
- Um Spark-Aktionen mit Apache Oozie zu verwenden, müssen Sie Ihrer Oozie-`workflow.xml`-Datei die folgende Konfiguration hinzufügen. Andernfalls fehlen mehrere wichtige Bibliotheken wie Hadoop und EMRFS im Klassenpfad der Spark-Executoren, die Oozie startet.

```
<spark-opts>--conf spark.yarn.populateHadoopClasspath=true</spark-opts>
```

- Wenn Sie Spark mit der Formatierung des Hive-Partitionsstandorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den Amazon-EMR-Versionen 5.30.0 bis 5.36.0 und 6.2.0 bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:
  - Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.
  - Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist `s3://bucket/table/p=a` ein Präfix von `s3://bucket/table/p=a b`.
  - Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen UTF-8-Wert, der kleiner als das `/`-Zeichen (U+002F) ist. Beispielsweise fällt das Leerzeichen (U+0020), das in `s3://bucket/table/p=a b` zwischen `a` und `b` vorkommt, in diese Kategorie. Beachten Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: `!"#$%&'()*+,-.` Weitere Informationen finden Sie unter [UTF-8-Kodierungstabelle und Unicode-Zeichen](#).

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.

## 6.4.0 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-*amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.4.1	Amazon SageMaker Spark-SDK
<code>emr-ddb</code>	4.16.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	3.2.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.5.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-notebook-env</code>	1.3.0	Conda Env für EMR-Notebooks, das Jupyter Enterprise Gateway enthält
<code>emr-s3-dist-cp</code>	2.18.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.

Komponente	Version	Beschreibung
emr-s3-select	2.1.0	EMR S3Select-Konnektor
emrfs	2,47,0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.13.1	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
flink-jobmanager-config	1.13.1	Verwaltung von Ressourcen auf EMR-Knoten für Apache JobManager Flink.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	3.2.1-amzn-4	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	3.2.1-amzn-4	HDFS-Service auf Knotenebene zum Speichern von Blöcken.

Komponente	Version	Beschreibung
hadoop-hdfs-library	3.2.1-amzn-4	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	3.2.1-amzn-4	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-hdfs-journalnode	3.2.1-amzn-4	HDFS-Service zum Verwalten des Hadoop-Dateisystem journals auf HA-Clustern.
hadoop-httpfs-server	3.2.1-amzn-4	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	3.2.1-amzn-4	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	3.2.1-amzn-4	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	3.2.1-amzn-4	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	3.2.1-amzn-4	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.



Komponente	Version	Beschreibung
hadoop-yarn-timeline-server	3.2.1-amzn-4	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	2.4.4-amzn-0	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	2.4.4-amzn-0	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	2.4.4-amzn-0	HBase-Befehlszeilen-Client.
hbase-rest-server	2.4.4-amzn-0	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	2.4.4-amzn-0	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	3.1.2-amzn-5	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	3.1.2-amzn-5	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	3.1.2-amzn-5	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.

Komponente	Version	Beschreibung
hive-client	3.1.2-amzn-5	Hive-Befehlszeilen-Client.
hive-hbase	3.1.2-amzn-5	Hive-hbase client.
hive-metastore-server	3.1.2-amzn-5	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	3.1.2-amzn-5	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.8.0-amzn-0	Inkrementelles Verarbeitungs-Framework zur Stromversorgung der Datenpipeline bei geringer Latenz und hoher Effizienz.
hudi-presto	0.8.0-amzn-0	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hudi-trino	0.8.0-amzn-0	Bündel-Bibliothek zum Ausführen von Trino mit Hudi.
hudi-spark	0.8.0-amzn-0	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.
hue-server	4.9.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	1.4.1	Multi-User-Server für Jupyter-Notebooks

Komponente	Version	Beschreibung
livy-server	0.7.1-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP- und Reverse-Proxy-Server.
mxnet	1.8.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.68+	MariaDB-Datenbankserver.
nvidia-cuda	10.1,243	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.2.1	Oozie-Befehlszeilen-Client.
oozie-server	5.2.1	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	5.1.2	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	5.1.2	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.254.1-amzn-0	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.

Komponente	Version	Beschreibung
presto-worker	0.254.1-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.254.1-amzn-0	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
trino-coordinator	359	Service zur Annahme von Abfragen und Verwaltung der Abfrageausführung der Trino-Worker.
trino-worker	359	Service für das Ausführen von Teilen einer Abfrage.
trino-client	359	Trino-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Trino-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	4,0.2	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	2.0.0	Apache Ranger Key Management System
spark-client	3.1.2-amzn-0	Spark-Befehlszeilen-Clients.

Komponente	Version	Beschreibung
spark-history-server	3.1.2-amzn-0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	3.1.2-amzn-0	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	3.1.2-amzn-0	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
spark-rapids	0.4.1	Nvidia-Spark-RAPIDS-Plugin, das Apache Spark mit GPUs beschleunigt.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	2.4.1	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.2	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.41+	Apache HTTP-Server.
zeppelin-server	0.9.0	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.

Komponente	Version	Beschreibung
zookeeper-server	3.5.7	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.5.7	ZooKeeper Befehlszeilen-Cli-ent.

## 6.4.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Aktionen zur Neukonfiguration treten auf, wenn Sie eine Konfiguration für Instance-Gruppen in einem laufenden Cluster angeben. Amazon EMR initiiert nur Rekonfigurationsaktionen für die Klassifizierungen, die Sie ändern. Weitere Informationen finden Sie unter [Eine Instance-Gruppe in einem laufenden Cluster neu konfigurieren](#).

### emr-6.4.0-Klassifizierungen

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.	Restarts the ResourceMananager service.
container-executor	Ändern Sie die Werte in der Datei „ <code>container-executor.cfg</code> “ Datei von Hadoop YARN.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
container-log4j	Ändert die Werte in der container-log4j.properties-Datei in Hadoop YARN.	Not available.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Ändern Sie die Docker-bezogenen Einstellungen.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
emrfs-site	Ändert die EMRFS-Einstellungen.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegistrator, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.	Restarts Flink history server.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.	Restarts Flink history server.
flink-log4j-session	Ändern Sie die Einstellungen von Flink log4j-session.properties für Kubernetes-/Yarn-Sitzungen.	Restarts Flink history server.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.	Restarts Flink history server.



Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.	Not available.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Ändert die Werte in der HBase-Umgebung.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.	Not available.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.	This classification should not be reconfigured.
hdfs-env	Ändert die Werte in der HDFS-Umgebung.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei proto-hive-site .xml von HCatalog.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.	Restarts Hive WebHCat server.
hive	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Hive.	Sets configurations to launch Hive LLAP service.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.	Not available.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.	Not available.
hive-env	Ändert die Werte in der Hive-Umgebung.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.	Not available.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log	Not available.
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.	Not available.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hue-ini	Ändert die Werte in der INI-Datei in Hue.	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.	Restarts Hadoop Httpfs service.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.	Not available.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.	Not available.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Ändern der Werte in der Hudi-Umgebung.	Not available.
hudi-defaults	Ändern Sie die Werte in der hudi-defaults.conf-Datei in Hudi.	Not available.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei <code>jupyterhub_config.py</code> .	Not available.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.	Not available.
jupyter-sparkmagic-conf	Ändert die Werte in der <code>config.json</code> -Datei in Sparkmagic.	Not available.
livy-conf	Ändert die Werte in der <code>livy.conf</code> -Datei von Livy.	Restarts Livy Server.
livy-env	Ändert die Werte in der Livy-Umgebung.	Restarts Livy Server.
livy-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Livy.	Restarts Livy Server.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Ändern Sie die Werte in der Datei <code>mapred-site.xml</code> der MapReduce Anwendung.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Ändert die Werte in der Oozie-Umgebung.	Restarts Oozie.
oozie-log4j	Ändert die Werte in der <code>oozie-log4j.properties</code> -Datei in Oozie.	Restarts Oozie.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.	Not available.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.	Not available.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.	Not available.
pig-env	Ändert die Werte in der Pig-Umgebung.	Not available.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.	Restarts Oozie.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.	Not available.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Ändert die Werte in der config.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.	Not available.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.	Not available.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.	Not available.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.	Not available.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.	Not available.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.	Not available.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.	Not available.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.	Not available.



Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.	Not available.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.	Not available.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.	Not available.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.	Not available.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.	Not available.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.	Not available.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.	Not available.
trino-log	Ändern Sie die Werte in der log.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-config	Ändern Sie die Werte in der config.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Ändern Sie die Werte in der Trino-Datei password-authenticator.properties.	Restarts Trino-Server (for Trino)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-env	Ändern Sie die Werte in der trino-env.sh-Datei von Trino.	Restarts Trino-Server (for Trino)
trino-node	Ändern Sie die Werte in der node.properties-Datei in Trino.	Not available.
trino-connector-blackhole	Ändern Sie die Werte in der blackhole.properties-Datei in Trino.	Not available.
trino-connector-cassandra	Ändern Sie die Werte in der cassandra.properties-Datei in Trino.	Not available.
trino-connector-hive	Ändern Sie die Werte in der hive.properties-Datei in Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Ändern Sie die Werte in der jmx.properties-Datei in Trino.	Not available.
trino-connector-kafka	Ändern Sie die Werte in der kafka.properties-Datei in Trino.	Not available.
trino-connector-localfile	Ändern Sie die Werte in der localfile.properties-Datei in Trino.	Not available.
trino-connector-memory	Ändern Sie die Werte in der memory.properties-Datei in Trino.	Not available.
trino-connector-mongodb	Ändern Sie die Werte in der mongodb.properties-Datei in Trino.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
trino-connector-mysql	Ändern Sie die Werte in der mysql.properties-Datei in Trino.	Not available.
trino-connector-postgresql	Ändern Sie die Werte in der postgresql.properties-Datei in Trino.	Not available.
trino-connector-raptor	Ändern Sie die Werte in der raptor.properties-Datei in Trino.	Not available.
trino-connector-redis	Ändern Sie die Werte in der redis.properties-Datei in Trino.	Not available.
trino-connector-redshift	Ändern Sie die Werte in der redshift.properties-Datei in Trino.	Not available.
trino-connector-tpch	Ändern Sie die Werte in der tpch.properties-Datei in Trino.	Not available.
trino-connector-tpcds	Ändern Sie die Werte in der tpcds.properties-Datei in Trino.	Not available.
ranger-kms-dbks-site	Ändert die Werte in der dbks-site.xml-Datei von Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Ändert die Werte in der Ranger KMS-Umgebung.	Restarts Ranger KMS Server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
ranger-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei von Ranger KMS.	Not available.
ranger-kms-db-ca	Ändert die Werte für die CA-Datei auf S3 für die MySQL SSL-Verbindung mit Ranger KMS.	Not available.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Ändert die Werte in der Spark-Umgebung.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.	Not available.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.	Not available.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.	Not available.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.	Restart Oozie and HiveServer2.
yarn-env	Ändert die Werte in der YARN-Umgebung.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.	Restarts Zeppelin.
zeppelin-site	Ändern Sie die Konfigurationseinstellungen in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.	Restarts Zookeeper server.

## Amazon-EMR-Version 6.3.1

### 6.3.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [PrestoSQL](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

#### Informationen zur Anwendungsversion

	emr-6.3.1	emr-6.3.0	emr-6.2.1	emr-6.2.0
AWS SDK for Java	1.11.977	1,11,977	1.11.880	1.11.880
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2.12,10	2.12.10	2.12.10	2.12.10
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.12.1	1.12.1	1.11.2	1.11.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2

	emr-6.3.1	emr-6.3.0	emr-6.2.1	emr-6.2.0
HBase	2.2.6	2.2.6	2.2.6-amzn-0	2.2.6-amzn-0
HCatalog	3.1.2	3.1.2	3.1.2	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.2	3.1.2	3.1.2	3.1.2
Hudi	0.7.0-amzn-0	0.7.0-amzn-0	0.6.0-amzn-1	0.6.0-amzn-1
Hue	4.9.0	4.9.0	4.8.0	4.8.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	2.1.0	2.1.0	2.1.0	2.1.0
JupyterHub	1.2.2	1.2.2	1.1.0	1.1.0
Livy	0.7.0	0.7.0	0.7.0	0.7.0
MXNet	1.7.0	1.7.0	1.7.0	1.7.0
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.0	5.2.0
Phoenix	5.0.0	5.0.0	5.0.0	5.0.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0,245,1	0,245,1	0,238,3	0,238,3
Spark	3.1.1	3.1.1	3.0.1	3.0.1
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.4.1	2.4.1	2.3.1	2.3.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2

	emr-6.3.1	emr-6.3.0	emr-6.2.1	emr-6.2.0
Trino (PrestoSQL)	350	350	343	343
Zeppelin	0.9.0	0.9.0	0.9.0	0.9.0
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

### 6.3.1 Versionshinweise

Dies ist eine Version zur Behebung von Problemen mit Amazon-EMR-Skalierung, wenn ein Cluster nicht erfolgreich hoch-/herunterskaliert werden kann oder Anwendungsfehler verursacht werden.

#### Änderungen, Verbesserungen und behobene Probleme

- Es wurde ein Problem behoben, bei dem Skalierungsanforderungen für einen großen, stark ausgelasteten Cluster fehlschlagen, wenn Amazon-EMR-On-Cluster-Daemons Aktivitäten zur Integritätsprüfung durchführten, z. B. das Erfassen des YARN-Knotenstatus und des HDFS-Knotenstatus. Das lag daran, dass Cluster-Daemons die Integritätsstatusdaten eines Knotens nicht an interne Amazon-EMR-Komponenten weitergeben konnten.
- Verbesserte EMR-On-Cluster-Daemons zur korrekten Nachverfolgung der Knotenstatus bei der Wiederverwendung von IP-Adressen, um die Zuverlässigkeit bei Skalierungsvorgängen zu verbessern.
- [SPARK-29683](#). Es wurde ein Problem behoben, bei dem während der Cluster-Skalierung Auftragsfehler auftraten, da Spark davon ausging, dass alle verfügbaren Knoten auf der Verweigern-Liste standen.
- [YARN-9011](#). Es wurde ein Problem behoben, bei dem Auftragsfehler aufgrund eines Fehlers bei der Außerbetriebnahme von YARN auftraten, wenn der Cluster versuchte, hoch- oder herunterzuskalieren.
- Das Problem mit Schritt- oder Auftragsfehlern bei der Cluster-Skalierung wurde behoben, indem sichergestellt wurde, dass die Knotenstatus zwischen den Amazon-EMR-On-Cluster-Daemons und YARN/HDFS immer konsistent sind.
- Es wurde ein Problem behoben, bei dem Clustervorgänge wie Herunterskalierung und Schrittübermittlung für Amazon-EMR-Cluster, die mit Kerberos-Authentifizierung aktiviert waren, fehlschlagen. Dies lag daran, dass der Amazon-EMR-On-Cluster-Daemon das Kerberos-Ticket



nicht erneuert hat, das für die sichere Kommunikation mit HDFS/YARN erforderlich ist, das auf dem Primärknoten ausgeführt wird.

- HTTPS ist jetzt standardmäßig für Amazon-Linux-Repositorys aktiviert. Wenn Sie eine Amazon-S3-VPCE-Richtlinie verwenden, um den Zugriff auf bestimmte Buckets zu beschränken, müssen Sie den neuen Amazon-Linux-Bucket-ARN `arn:aws:s3:::amazonlinux-2-repos-$region/*` zu Ihrer Richtlinie hinzufügen (ersetzen Sie `$region` durch die Region, in der sich der Endpunkt befindet). Weitere Informationen finden Sie unter diesem Thema in den AWS Diskussionsforen. [Ankündigung: Amazon Linux 2 unterstützt jetzt die Möglichkeit, HTTPS zu verwenden, wenn eine Verbindung zu Paket-Repositorys hergestellt wird.](#)

## Bekannte Probleme

- Wenn Sie Spark mit der Formatierung des Hive-Partitionsstandorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den Amazon-EMR-Versionen 5.30.0 bis 5.36.0 und 6.2.0 bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:
  - Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.
  - Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist `s3://bucket/table/p=a` ein Präfix von `s3://bucket/table/p=a b`.
  - Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen UTF-8-Wert, der kleiner als das `/`-Zeichen (U+002F) ist. Beispielsweise fällt das Leerzeichen (U+0020), das in `s3://bucket/table/p=a b` zwischen `a` und `b` vorkommt, in diese Kategorie. Beachten Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: `!"#$%&'()*+,-.` Weitere Informationen finden Sie unter [UTF-8-Kodierungstabelle und Unicode-Zeichen](#).

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.

### 6.3.1 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version

sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-amzn-*EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen myapp-component der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als 2.2-amzn-2 aufgeführt.

Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1.4.1	Amazon SageMaker Spark-SDK
emr-ddb	4.16.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	3.2.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.5.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-notebook-env	1.2.0	Conda Env für EMR-Notebooks, das Jupyter Enterprise Gateway enthält
emr-s3-dist-cp	2.18.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	2.1.0	EMR S3Select-Konnektor
emrfs	2,46,0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
flink-client	1.12.1	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
flink-jobmanager-config	1.12.1	Verwaltung von Ressourcen auf EMR-Knoten für Apache JobManager Flink.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	3.2.1-amzn-3.1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	3.2.1-amzn-3.1	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	3.2.1-amzn-3.1	HDFS-Client und -Bibliothek für die Befehlszeile

Komponente	Version	Beschreibung
hadoop-hdfs-namenode	3.2.1-amzn-3.1	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-hdfs-journalnode	3.2.1-amzn-3.1	HDFS-Service zum Verwalten des Hadoop-Dateisystem journals auf HA-Clustern.
hadoop-httpfs-server	3.2.1-amzn-3.1	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	3.2.1-amzn-3.1	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	3.2.1-amzn-3.1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	3.2.1-amzn-3.1	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	3.2.1-amzn-3.1	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	3.2.1-amzn-3.1	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.

Komponente	Version	Beschreibung
hbase-hmaster	2.2.6-amzn-1	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	2.2.6-amzn-1	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	2.2.6-amzn-1	HBase-Befehlszeilen-Client.
hbase-rest-server	2.2.6-amzn-1	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	2.2.6-amzn-1	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	3.1.2-amzn-4	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	3.1.2-amzn-4	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	3.1.2-amzn-4	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	3.1.2-amzn-4	Hive-Befehlszeilen-Client.
hive-hbase	3.1.2-amzn-4	Hive-hbase client.

Komponente	Version	Beschreibung
hive-metastore-server	3.1.2-amzn-4	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	3.1.2-amzn-4	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.7.0-amzn-0	Inkrementelles Verarbeitungs-Framework zur Stromversorgung der Datenpipeline bei geringer Latenz und hoher Effizienz.
hudi-presto	0.7.0-amzn-0	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hudi-prestosql	0.7.0-amzn-0	Bündel-Bibliothek zum Ausführen von PrestoSQL mit Hudi.
hudi-spark	0.7.0-amzn-0	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.
hue-server	4.9.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	1.2.2	Multi-User-Server für Jupyter-Notebooks
livy-server	0.7.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark

Komponente	Version	Beschreibung
nginx	1.12.1	nginx [engine x] ist ein HTTP- und Reverse-Proxy-Server.
mxnet	1.7.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.68+	MariaDB-Datenbankserver.
nvidia-cuda	10.1,243	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.2.1	Oozie-Befehlszeilen-Client.
oozie-server	5.2.1	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	5.0.0-HBase-2.0	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	5.0.0-HBase-2.0	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.245.1-amzn-0	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.245.1-amzn-0	Service für das Ausführen von Teilen einer Abfrage.

Komponente	Version	Beschreibung
presto-client	0.245.1-amzn-0	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
prestoql-coordinator	350	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Prestoql-Worker.
prestoql-worker	350	Service für das Ausführen von Teilen einer Abfrage.
prestoql-client	350	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	4.0.2	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	2.0.0	Apache Ranger Key Management System
spark-client	3.1.1-amzn-0.1	Spark-Befehlszeilen-Clients.
spark-history-server	3.1.1-amzn-0.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.



Komponente	Version	Beschreibung
spark-on-yarn	3.1.1-amzn-0.1	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	3.1.1-amzn-0.1	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
spark-rapids	0.4.1	Nvidia-Spark-RAPIDS-Plugin, das Apache Spark mit GPUs beschleunigt.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	2.4.1	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.2	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.41+	Apache HTTP-Server.
zeppelin-server	0.9.0	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.14	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.

Komponente	Version	Beschreibung
zookeeper-client	3.4.14	ZooKeeper Befehlszeilen-Client.

### 6.3.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Aktionen zur Neukonfiguration treten auf, wenn Sie eine Konfiguration für Instance-Gruppen in einem laufenden Cluster angeben. Amazon EMR initiiert nur Rekonfigurationsaktionen für die Klassifizierungen, die Sie ändern. Weitere Informationen finden Sie unter [Eine Instance-Gruppe in einem laufenden Cluster neu konfigurieren](#).

#### emr-6.3.1-Klassifizierungen

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.	Restarts the ResourceManager service.
container-executor	Ändern Sie die Werte in der Datei „ <code>container-executor.cfg</code> “ Datei von Hadoop YARN.	Not available.
container-log4j	Ändert die Werte in der <code>container-log4j.properties</code> -Datei in Hadoop YARN.	Not available.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager,

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
		NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Ändern Sie die Docker-bezogenen Einstellungen.	Not available.
emrfs-site	Ändert die EMRFS-Einstellungen.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegionserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.	Restarts Flink history server.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.	Restarts Flink history server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
flink-log4j-session	Ändern Sie die Einstellungen von Flink log4j-session.properties für Kubernetes-/Yarn-Sitzungen.	Restarts Flink history server.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.	Restarts Flink history server.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.	Not available.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.	Not available.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Ändert die Werte in der HBase-Umgebung.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.	Not available.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.	This classification should not be reconfigured.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hdfs-env	Ändert die Werte in der HDFS-Umgebung.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei proto-hive-site .xml von HCatalog.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.	Restarts Hive WebHCat server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hive	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.	Not available.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.	Not available.
hive-env	Ändert die Werte in der Hive-Umgebung.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.	Not available.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log	Not available.
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.	Not available.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.	Not available.
hue-ini	Ändert die Werte in der INI-Datei in Hue.	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.	Restarts Hadoop Httpfs service.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.	Not available.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.	Not available.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Ändern der Werte in der Hudi-Umgebung.	Not available.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.	Not available.



Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei <code>jupyterhub_config.py</code> .	Not available.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.	Not available.
jupyter-sparkmagic-conf	Ändert die Werte in der <code>config.json</code> -Datei in Sparkmagic.	Not available.
livy-conf	Ändert die Werte in der <code>livy.conf</code> -Datei von Livy.	Restarts Livy Server.
livy-env	Ändert die Werte in der Livy-Umgebung.	Restarts Livy Server.
livy-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Livy.	Restarts Livy Server.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Ändern Sie die Werte in der Datei <code>mapred-site.xml</code> der MapReduce Anwendung.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Ändert die Werte in der Oozie-Umgebung.	Restarts Oozie.
oozie-log4j	Ändert die Werte in der <code>oozie-log4j.properties</code> -Datei in Oozie.	Restarts Oozie.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.	Not available.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.	Not available.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.	Not available.
pig-env	Ändert die Werte in der Pig-Umgebung.	Not available.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.	Restarts Oozie.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.	Not available.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Ändert die Werte in der config.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.	Not available.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.	Not available.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.	Not available.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.	Not available.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.	Not available.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.	Not available.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.	Not available.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.	Not available.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.	Not available.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.	Not available.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.	Not available.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.	Not available.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.	Not available.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.	Not available.
prestosql-log	Ändert die Werte in der log.properties-Datei in Presto.	Restarts Presto-Server (for PrestoSQL)
prestosql-config	Ändert die Werte in der config.properties-Datei in Presto.	Restarts Presto-Server (for PrestoSQL)
prestosql-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.	Restarts Presto-Server (for PrestoSQL)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
prestosql-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.	Restarts Presto-Server (for PrestoSQL)
prestosql-node	Ändern Sie die Werte in PrestoSQLs node.properties-Datei.	Not available.
prestosql-connector-blackhole	Ändern Sie die Werte in PrestoSQLs blackhole.properties-Datei.	Not available.
prestosql-connector-cassandra	Ändern Sie die Werte in PrestoSQLs cassandra.properties-Datei.	Not available.
prestosql-connector-hive	Ändern Sie die Werte in PrestoSQLs hive.properties-Datei.	Restarts Presto-Server (for PrestoSQL)
prestosql-connector-jmx	Ändern Sie die Werte in PrestoSQLs jmx.properties-Datei.	Not available.
prestosql-connector-kafka	Ändern Sie die Werte in PrestoSQLs kafka.properties-Datei.	Not available.
prestosql-connector-localfile	Ändern Sie die Werte in PrestoSQLs localfile.properties-Datei.	Not available.
prestosql-connector-memory	Ändern Sie die Werte in PrestoSQLs memory.properties-Datei in Presto.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
prestosql-connector-mongodb	Ändern Sie die Werte in PrestoSQLs mongod.properties-Datei in Presto.	Not available.
prestosql-connector-mysql	Ändern Sie die Werte in PrestoSQLs mysql.properties-Datei.	Not available.
prestosql-connector-postgresql	Ändern Sie die Werte in PrestoSQLs postgresql.properties-Datei.	Not available.
prestosql-connector-raptor	Ändern Sie die Werte in PrestoSQLs raptor.properties-Datei.	Not available.
prestosql-connector-redis	Ändern Sie die Werte in PrestoSQLs redis.properties-Datei.	Not available.
prestosql-connector-redshift	Ändern Sie die Werte in PrestoSQLs redshift.properties-Datei.	Not available.
prestosql-connector-tpch	Ändern Sie die Werte in PrestoSQLs tpch.properties-Datei.	Not available.
prestosql-connector-tpcds	Ändern Sie die Werte in PrestoSQLs tpcds.properties-Datei.	Not available.
ranger-kms-dbks-site	Ändert die Werte in der dbks-site.xml-Datei von Ranger KMS.	Restarts Ranger KMS Server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Ändert die Werte in der Ranger KMS-Umgebung.	Restarts Ranger KMS Server.
ranger-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei von Ranger KMS.	Not available.
ranger-kms-db-ca	Ändert die Werte für die CA-Datei auf S3 für die MySQL SSL-Verbindung mit Ranger KMS.	Not available.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Ändert die Werte in der Spark-Umgebung.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.	Not available.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.	Not available.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.	Not available.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.	Not available.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.	Restart Oozie and HiveServer2.
yarn-env	Ändert die Werte in der YARN-Umgebung.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.	Restarts Zeppelin.
zeppelin-site	Ändern Sie die Konfigurationseinstellungen in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.	Restarts Zookeeper server.



Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.	Restarts Zookeeper server.

## Amazon-EMR-Version 6.3.0

### 6.3.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [PrestoSQL](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

### Informationen zur Anwendungsversion

	emr-6.3.0	emr-6.2.1	emr-6.2.0	emr-6.1.1
AWS SDK for Java	1.11.977	1.11.880	1.11.880	1.11.828
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2.12,10	2.12.10	2.12.10	2.12.10

	emr-6.3.0	emr-6.2.1	emr-6.2.0	emr-6.1.1
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.12.1	1.11.2	1.11.2	1.11.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.2.6	2.2.6-amzn-0	2.2.6-amzn-0	2.2.5
HCatalog	3.1.2	3.1.2	3.1.2	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.2	3.1.2	3.1.2	3.1.2
Hudi	0.7.0-amzn-0	0.6.0-amzn-1	0.6.0-amzn-1	0.5.2-incubating-amzn-2
Hue	4.9.0	4.8.0	4.8.0	4.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	2.1.0	2.1.0	2.1.0	-
JupyterHub	1.2.2	1.1.0	1.1.0	1.1.0
Livy	0.7.0	0.7.0	0.7.0	0.7.0
MXNet	1.7.0	1.7.0	1.7.0	1.6.0
Mahout	-	-	-	-
Oozie	5.2.1	5.2.0	5.2.0	5.2.0
Phoenix	5.0.0	5.0.0	5.0.0	5.0.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0

	emr-6.3.0	emr-6.2.1	emr-6.2.0	emr-6.1.1
Presto	0,245,1	0,238,3	0,238,3	0.232
Spark	3.1.1	3.0.1	3.0.1	3.0.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.4.1	2.3.1	2.3.1	2.1.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	350	343	343	338
Zeppelin	0.9.0	0.9.0	0.9.0	0.9.0
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

## 6.3.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 6.3.0. Änderungen beziehen sich auf Version 6.2.0.

Erste Version: 12. Mai 2021

Letzte Aktualisierung: 9. August 2021

### Unterstützte Anwendungen

- AWS SDK for Java Ausführung 1.11.977
- CloudWatch Version 2.1.0 senken
- Version 4.16.0 von EMR DynamoDB Connector
- EMRFS-Version 2.46.0
- Version 3.2.0 von Amazon EMR Goodies
- Version 3.5.0 von Amazon EMR Kinesis Connector
- Version 2.0.0 von Amazon EMR Record Server
- Version 2.5.0 von Amazon EMR Skripts

- Flink-Version 1.12.1
- Ganglia Version 3.7.2
- AWS Glue Hive Metastore Client Version 3.2.0
- Hadoop-Version 3.2.1-amzn-3
- HBase-Version 2.2.6-amzn-1
- ase-operator-tools HB 1.0.0
- HCatalog-Version 3.1.2-amzn-0
- Hive-Version 3.1.2-amzn-4
- Hudi-Version 0.7.0-amzn-0
- Hue Version 4.9.0
- Java-JDK-Version Corretto-8.282.08.1 (Build 1.8.0\_282-b08)
- JupyterHub Ausführung 1.2.0
- Livy-Version 0.7.0-incubating
- MXNet-Version 1.7.0
- Oozie-Version 5.2.1
- Phoenix Version 5.0.0
- Pig Version 0.17.0
- Presto-Version 0.245.1-amzn-0
- PrestoSQL-Version 350
- Version 2.0.0 von Apache Ranger KMS (transparente Multi-Master-Verschlüsselung)
- Ranger-Plugins 2.0.1-amzn-0
- Ranger-S3-Plugin 1.1.0
- SageMaker Spark SDK, Version 1.4.1
- Scala-Version 2.12.10 (OpenJDK 64-Bit-Server-VM, Java 1.8.0\_282)
- Spark-Version 3.1.1-amzn-0
- Spark-Rapids 0.4.1
- Sqoop-Version 1.4.7
- TensorFlow Ausführung 2.4.1
- Tez-Version 0.9.2

- Zeppelin-Version 0.9.0
- Zookeeper Version 3.4.14
- Konnektoren und Treiber: DynamoDB Connector 4.16.0

## Neue Features

- Amazon EMR unterstützt Amazon S3 Access Points, ein Feature von Amazon S3, mit dem Sie den Zugriff auf gemeinsam genutzte Data Lakes einfach verwalten können. Mit Ihrem Alias von Amazon S3 Access Point können Sie Ihren Datenzugriff auf Amazon EMR in großem Umfang vereinfachen. Sie können Amazon S3 Access Points mit allen Versionen von Amazon EMR ohne zusätzliche Kosten in allen AWS Regionen verwenden, in denen Amazon EMR verfügbar ist. Weitere Informationen zu Amazon S3 Access Points und Zugangspunkt-Aliassen finden Sie unter [Verwenden eines Alias im Bucket-Stil für Ihren Zugangspunkt](#) im Amazon-S3-Benutzerhandbuch.
- Neue `DescribeReleaseLabel` und `ListReleaseLabel` API-Parameter bieten Details zum Amazon-EMR-Release-Label. Sie können programmgesteuert Versionen auflisten, die in der Region verfügbar sind, in der die API-Anfrage ausgeführt wird, und die verfügbaren Anwendungen für ein bestimmtes Amazon-EMR-Release-Label auflisten. In den Release-Label-Parametern sind auch Amazon-EMR-Versionen aufgeführt, die eine bestimmte Anwendung unterstützen, z. B. Spark. Diese Informationen können verwendet werden, um Amazon-EMR-Cluster programmgesteuert zu starten. Sie können beispielsweise einen Cluster mit der neuesten Release-Version aus den `ListReleaseLabel`-Ergebnissen starten. Weitere Informationen finden Sie unter [DescribeReleaseLabel](#) und [ListReleaseLabels](#) in der Amazon EMR API-Referenz.
- Mit Amazon EMR 6.3.0 können Sie einen Cluster starten, der nativ in Apache Ranger integriert ist. Apache Ranger ist ein Open-Source-Framework zur Aktivierung, Überwachung und Verwaltung einer umfassenden Datensicherheit auf der gesamten Hadoop-Plattform. Weitere Informationen finden Sie unter [Apache Ranger](#). Dank der nativen Integration können Sie Ihren eigenen Apache Ranger verwenden, um eine detaillierte Datenzugriffskontrolle auf Amazon EMR durchzusetzen. Weitere Informationen [finden Sie unter Amazon EMR mit Apache Ranger](#) im Verwaltungshandbuch für Amazon EMR.
- Verwaltete Richtlinien mit Geltungsbereich: Um den AWS bewährten Methoden zu entsprechen, hat Amazon EMR verwaltete Standardrichtlinien mit EMR-Geltungsbereich der Version v2 eingeführt, die als Ersatz für Richtlinien dienen, die nicht mehr unterstützt werden. Weitere Informationen finden Sie unter [Verwaltete Richtlinien von Amazon EMR](#).
- Unterstützungsstatus des Instance Metadata Service (IMDS) V2: Für Amazon EMR 6.2 oder höher verwenden Amazon-EMR-Komponenten IMDSv2 für alle IMDS-Aufrufe. Für IMDS-Aufrufe in

Ihrem Anwendungscode können Sie sowohl IMDSv1 als auch IMDSv2 verwenden oder das IMDS so konfigurieren, dass es aus Sicherheitsgründen nur IMDSv2 verwendet. Wenn Sie IMDSv1 in früheren Amazon-EMR-6.x-Versionen deaktivieren, führt dies zu einem Cluster-Startup-Fehler.

## Änderungen, Verbesserungen und behobene Probleme

- Dies ist eine Version zur Behebung von Problemen mit Amazon-EMR-Skalierung, wenn ein Cluster nicht erfolgreich hoch-/herunterskaliert werden kann oder Anwendungsfehler verursacht werden.
- Es wurde ein Problem behoben, bei dem Skalierungsanforderungen für einen großen, stark ausgelasteten Cluster fehlschlagen, wenn Amazon-EMR-On-Cluster-Daemons Aktivitäten zur Integritätsprüfung durchführten, z. B. das Erfassen des YARN-Knotenstatus und des HDFS-Knotenstatus. Das lag daran, dass Cluster-Daemons die Integritätsstatusdaten eines Knotens nicht an interne Amazon-EMR-Komponenten weitergeben konnten.
- Verbesserte EMR-On-Cluster-Daemons zur korrekten Nachverfolgung der Knotenstatus bei der Wiederverwendung von IP-Adressen, um die Zuverlässigkeit bei Skalierungsvorgängen zu verbessern.
- [SPARK-29683](#). Es wurde ein Problem behoben, bei dem während der Cluster-Skalierung Auftragsfehler auftraten, da Spark davon ausging, dass alle verfügbaren Knoten auf der Verweigern-Liste standen.
- [YARN-9011](#). Es wurde ein Problem behoben, bei dem Auftragsfehler aufgrund eines Fehlers bei der Außerbetriebnahme von YARN auftraten, wenn der Cluster versuchte, hoch- oder herunterzuskalieren.
- Das Problem mit Schritt- oder Auftragsfehlern bei der Cluster-Skalierung wurde behoben, indem sichergestellt wurde, dass die Knotenstatus zwischen den Amazon-EMR-On-Cluster-Daemons und YARN/HDFS immer konsistent sind.
- Es wurde ein Problem behoben, bei dem Clustervorgänge wie Herunterskalierung und Schrittübermittlung für Amazon-EMR-Cluster, die mit Kerberos-Authentifizierung aktiviert waren, fehlschlagen. Dies lag daran, dass der Amazon-EMR-On-Cluster-Daemon das Kerberos-Ticket nicht erneuert hat, das für die sichere Kommunikation mit HDFS/YARN erforderlich ist, das auf dem Primärknoten ausgeführt wird.
- Neuere Amazon-EMR-Versionen beheben das Problem mit einem niedrigeren Limit für die „maximale Anzahl geöffneter Dateien“ auf älteren AL2-Versionen in Amazon EMR. Die Amazon-EMR-Versionen 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 und höher enthalten jetzt einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“.

- Der Erklärungsmodus der Spark-SQL-Benutzeroberfläche wurde in [Spark 3.1](#) von `extended` auf `formatted` geändert. Amazon EMR hat ihn wieder auf `extended` zurückgesetzt, um logische Planinformationen in die Spark-SQL-Benutzeroberfläche aufzunehmen. Dies kann durch Einstellen von `spark.sql.ui.explainMode` auf `formatted` rückgängig gemacht werden.
- Die folgenden Commits wurden aus dem Spark-Master-Branch zurückportiert.
  - [\[SPARK-34752\]](#) [\[BUILD\]](#) Bump Jetty auf 9.4.37, um CVE-2020-27223 zu adressieren.
  - [\[SPARK-34534\]](#) Korrigiert die Reihenfolge der BlockIDs beim Abrufen von Blöcken.  
`FetchShuffleBlocks`
  - [\[SPARK-34681\]](#) [\[SQL\]](#) Es wurde ein Fehler behoben, der beim Erstellen der linken Seite mit einer ungleichen Bedingung beim Erstellen einer linken Seite auftrat.
  - [\[SPARK-34497\]](#) [\[SQL\]](#) Korrigiert integrierte JDBC-Verbindungsanbieter, um Änderungen im JVM-Sicherheitskontext wiederherzustellen.
- Um die Interoperabilität mit dem Nvidia Spark RAPIDS-Plugin zu verbessern, wurde eine Problemumgehung hinzugefügt, um ein Problem zu beheben, das verhindert, dass dynamisches Partitionsbereinigen ausgelöst wird, wenn Nvidia Spark RAPIDS mit deaktivierter adaptiver Abfrageausführung verwendet wird. Weitere Informationen finden Sie unter [RAPIDS-Problem #1378](#) und [RAPIDS-Problem Nr. #1386](#). Einzelheiten zur neuen `spark.sql.optimizer.dynamicPartitionPruning.enforceBroadcastReuse`-Konfiguration finden Sie in der [RAPIDS-Problem ##1386](#).
- Der Standardalgorithmus für den Dateiausgabe-Committer wurde in Open Source Spark 3.1 vom v2-Algorithmus auf den v1-Algorithmus geändert. Weitere Informationen finden Sie unter [Amazon EMR zur Optimierung der Spark-Leistung – dynamisches Bereinigen von Partitionen](#).
- Amazon EMR hat auf den v2-Algorithmus zurückgesetzt, den in früheren Amazon-EMR-6.x-Versionen verwendeten Standard, um eine Leistungsregression zu verhindern. Um das Open-Source-Verhalten von Spark 3.1 wiederherzustellen, setzen Sie `spark.hadoop.mapreduce.fileoutputcommitter.algorithm.version` auf 1. Open Source Spark hat diese Änderung vorgenommen, weil das Festschreiben von Aufgaben im Dateioutput-Committer-Algorithmus v2 nicht atomar ist, was in einigen Fällen zu Problemen mit der Korrektheit der Ausgabedaten führen kann. Das Festschreiben von Aufgaben im Algorithmus v1 ist jedoch auch nicht atomar. In einigen Szenarien beinhaltet die Aufgabenübergabe ein Löschen, das vor einer Umbenennung durchgeführt wird. Dies kann zu einem unbemerkten Problem mit der Datenkorrektheit führen.

- Probleme bei der verwalteten Skalierung in früheren Amazon-EMR-Versionen wurden behoben und Verbesserungen vorgenommen, sodass die Ausfallraten von Anwendungen erheblich reduziert wurden.
- Das AWS Java SDK-Bundle wurde auf jedem neuen Cluster installiert. Dies ist ein einzelnes JAR, das alle Service-SDKs und ihre Abhängigkeiten enthält, anstatt einzelne Komponenten-Jars. Weitere Informationen finden Sie unter [Gebündelte Java SDK Abhängigkeit](#)

## Bekannte Probleme

- Für private Subnetz-Cluster von Amazon EMR 6.3.0 und 6.2.0 können Sie nicht auf die Ganglia-Webbenutzeroberfläche zugreifen. Sie erhalten die Fehlermeldung „Zugriff verweigert (403)“. Andere Web-Benutzeroberflächen wie Spark, Hue JupyterHub, Zeppelin, Livy und Tez funktionieren normal. Der Zugriff auf die Ganglia-Web-Benutzeroberfläche auf öffentlichen Subnetzclustern funktioniert ebenfalls normal. Um dieses Problem zu beheben, starten Sie den httpd-Service auf dem Primärknoten mit `sudo systemctl restart httpd` neu. Dieses Problem wurde in Amazon EMR 6.4.0 behoben.
- Wenn der AWS Glue-Datenkatalog aktiviert ist, kann die Verwendung von Spark für den Zugriff auf eine AWS Glue-Datenbank mit einer Null-Ort-URI fehlschlagen. Dies ist bei früheren Amazon-EMR-Versionen der Fall, aber mit SPARK-31709 (<https://issues.apache.org/jira/browse/SPARK-31709>) gilt es für mehr Fälle. Wenn Sie beispielsweise eine Tabelle in der AWS Standard-Glue-DB erstellen, deren Standort-URI eine Nullzeichenfolge ist, `spark.sql("CREATE TABLE mytest (key string) location '/table_path';")` schlägt dies mit der Meldung „Cannot create a Path from an empty string“ fehl. Um dies zu umgehen, legen Sie manuell einen Standort-URI Ihrer AWS Glue-Datenbanken fest und erstellen Sie dann mit Spark Tabellen innerhalb dieser Datenbanken.
- In Amazon EMR 6.3.0 wurde PrestoSQL von Version 343 auf Version 350 aktualisiert. Es gibt zwei sicherheitsrelevante Änderungen von Open Source, die sich auf diese Versionsänderung beziehen. Die dateibasierte Katalogzugriffssteuerung wurde von `deny` zu `allow` geändert, wenn keine Regeln für Tabellen-, Schema- oder Sitzungseigenschaften definiert sind. Außerdem wurde die dateibasierte Systemzugriffskontrolle dahingehend geändert, dass sie Dateien unterstützt, für die keine Katalogregeln definiert sind. In diesem Fall ist der gesamte Zugriff auf Kataloge zulässig.

Weitere Informationen finden Sie in [Version 344 \(9. Oktober 2020\)](#).

- Beachten Sie, dass das Hadoop-Benutzerverzeichnis (`/home/hadoop`) für jeden lesbar ist. Es verfügt über Unix 755 (`drwxr-xr-x`)-Verzeichnisberechtigungen, um den Lesezugriff durch Frameworks wie Hive zu ermöglichen. Sie können Dateien in `/home/hadoop` und seinen



Unterverzeichnissen ablegen, beachten Sie jedoch die Berechtigungen für diese Verzeichnisse, um vertrauliche Informationen zu schützen.

- Niedrigeres Limit für die „Max. Anzahl geöffneter Dateien“ auf älteren AL2-Versionen [in neueren Versionen behoben]. Amazon-EMR-Versionen: emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0 und emr-6.2.0 basieren auf älteren Versionen von Amazon Linux 2 (AL2), die eine niedrigere Ulimit-Einstellung für „Max. Anzahl geöffneter Dateien“ haben, wenn Amazon-EMR-Cluster mit dem Standard-AMI erstellt werden. Die Amazon-EMR-Versionen 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 und höher enthalten einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“. Versionen mit einem niedrigeren Limit für geöffnete Dateien verursachen beim Senden des Spark-Jobs den Fehler „Zu viele offene Dateien“. In den betroffenen Versionen hat das Amazon-EMR-Standard-AMI eine Ulimit-StandardEinstellung von 4 096 für „Max. Anzahl geöffneter Dateien“, was unter dem Dateilimit von 65 536 im neuesten Amazon-Linux-2-AMI liegt. Die niedrigere Ulimit-Einstellung für „Max. Anzahl geöffneter Dateien“ führt dazu, dass der Spark-Job fehlschlägt, wenn der Spark-Treiber und der Executor versuchen, mehr als 4 096 Dateien zu öffnen. Um das Problem zu beheben, verfügt Amazon EMR über ein Bootstrap Action (BA)-Skript, das die Ulimit-Einstellung bei der Cluster-Erstellung anpasst.

Wenn Sie eine ältere Amazon-EMR-Version verwenden, für die dieses Problem nicht dauerhaft behoben ist, können Sie mit der folgenden Problemumgehung das Instance-Controller-Ulimit explizit auf maximal 65 536 Dateien festlegen.

Explizit ein ulimit über die Befehlszeile setzen

1. Bearbeiten Sie `/etc/systemd/system/instance-controller.service`, um die folgenden Parameter zum Abschnitt Service hinzuzufügen.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Starten Sie neu InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Mithilfe der Bootstrap-Aktion (BA) ein Ulimit festlegen

Sie können auch ein Bootstrap-Aktionsskript (BA) verwenden, um das Ulimit für den Instance-Controller bei der Clustererstellung auf 65 536 Dateien zu konfigurieren.

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

**⚠ Important**

EMR-Cluster, auf denen Amazon-Linux- oder Amazon-Linux-2-AMIs (Amazon Machine Images) ausgeführt werden, verwenden das Standardverhalten von Amazon Linux und laden wichtige und kritische Kernel-Updates, die einen Neustart erfordern, nicht automatisch herunter und installieren sie. Dies ist dasselbe Verhalten wie bei anderen Amazon-EC2-Instances, die das standardmäßige Amazon-Linux-AMI ausführen. Wenn neue Amazon-Linux-Softwareupdates, die einen Neustart erfordern (wie Kernel-, NVIDIA- und CUDA-Updates), nach der Veröffentlichung einer Amazon-EMR-Version verfügbar werden, laden EMR-Cluster-Instances, die das Standard-AMI ausführen, diese Updates nicht automatisch herunter und installieren sie. Um Kernel-Updates zu erhalten, können Sie [Ihr Amazon-EMR-AMI](#) so anpassen, dass es [das neueste Amazon-Linux-AMI verwendet](#).

- Um Spark-Aktionen mit Apache Oozie zu verwenden, müssen Sie Ihrer Oozie-workflow.xml-Datei die folgende Konfiguration hinzufügen. Andernfalls fehlen mehrere wichtige Bibliotheken wie Hadoop und EMRFS im Klassenpfad der Spark-Executoren, die Oozie startet.

```
<spark-opts>--conf spark.yarn.populateHadoopClasspath=true</spark-opts>
```

- Wenn Sie Spark mit der Formatierung des Hive-Partitionsstandorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den Amazon-EMR-Versionen 5.30.0 bis 5.36.0 und 6.2.0 bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:
  - Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.
  - Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist `s3://bucket/table/p=a` ein Präfix von `s3://bucket/table/p=a b`.
  - Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen UTF-8-Wert, der kleiner als das `/`-Zeichen (U+002F) ist. Beispielsweise fällt das Leerzeichen (U+0020), das in `s3://bucket/table/p=a b` zwischen `a` und `b` vorkommt, in diese Kategorie. Beachten Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: `!"#$%&'()*+,-.` Weitere Informationen finden Sie unter [UTF-8-Kodierungstabelle und Unicode-Zeichen](#).

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.

### 6.3.0 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1.4.1	Amazon SageMaker Spark-SDK
emr-ddb	4.16.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	3.2.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.5.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-notebook-env	1.2.0	Conda Env für EMR-Notebooks, das Jupyter Enterprise Gateway enthält
emr-s3-dist-cp	2.18.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	2.1.0	EMR S3Select-Konnektor
emrfs	2,46,0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.12.1	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
flink-jobmanager-config	1.12.1	Verwaltung von Ressourcen auf EMR-Knoten für Apache JobManager Flink.

Komponente	Version	Beschreibung
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	3.2.1-amzn-3	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	3.2.1-amzn-3	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	3.2.1-amzn-3	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	3.2.1-amzn-3	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-hdfs-journalnode	3.2.1-amzn-3	HDFS-Service zum Verwalten des Hadoop-Dateisystem journals auf HA-Clustern.

Komponente	Version	Beschreibung
hadoop-httfs-server	3.2.1-amzn-3	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	3.2.1-amzn-3	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	3.2.1-amzn-3	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	3.2.1-amzn-3	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	3.2.1-amzn-3	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	3.2.1-amzn-3	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	2.2.6-amzn-1	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	2.2.6-amzn-1	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.

Komponente	Version	Beschreibung
hbase-client	2.2.6-amzn-1	HBase-Befehlszeilen-Client.
hbase-rest-server	2.2.6-amzn-1	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	2.2.6-amzn-1	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	3.1.2-amzn-4	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	3.1.2-amzn-4	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	3.1.2-amzn-4	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	3.1.2-amzn-4	Hive-Befehlszeilen-Client.
hive-hbase	3.1.2-amzn-4	Hive-hbase client.
hive-metastore-server	3.1.2-amzn-4	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	3.1.2-amzn-4	Service zur Annahme von Hive-Abfragen als Webanfragen.

Komponente	Version	Beschreibung
hudi	0.7.0-amzn-0	Inkrementelles Verarbeitungs-Framework zur Stromversorgung der Datenpipeline bei geringer Latenz und hoher Effizienz.
hudi-presto	0.7.0-amzn-0	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hudi-prestosql	0.7.0-amzn-0	Bündel-Bibliothek zum Ausführen von PrestoSQL mit Hudi.
hudi-spark	0.7.0-amzn-0	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.
hue-server	4.9.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	1.2.2	Multi-User-Server für Jupyter-Notebooks
livy-server	0.7.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP- und Reverse-Proxy-Server.
mxnet	1.7.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.68+	MariaDB-Datenbankserver.



Komponente	Version	Beschreibung
nvidia-cuda	10.1,243	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.2.1	Oozie-Befehlszeilen-Client.
oozie-server	5.2.1	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	5.0.0-HBase-2.0	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	5.0.0-HBase-2.0	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.245.1-amzn-0	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.245.1-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.245.1-amzn-0	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.

Komponente	Version	Beschreibung
prestoql-coordinator	350	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Prestoql-Worker.
prestoql-worker	350	Service für das Ausführen von Teilen einer Abfrage.
prestoql-client	350	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	4.0.2	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	2.0.0	Apache Ranger Key Management System
spark-client	3.1.1-amzn-0	Spark-Befehlszeilen-Clients.
spark-history-server	3.1.1-amzn-0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	3.1.1-amzn-0	In-Memory-Ausführungs-Engine für YARN.

Komponente	Version	Beschreibung
spark-yarn-slave	3.1.1-amzn-0	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
spark-rapids	0.4.1	Nvidia-Spark-RAPIDS-Plugin, das Apache Spark mit GPUs beschleunigt.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	2.4.1	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.2	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.41+	Apache HTTP-Server.
zeppelin-server	0.9.0	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.14	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.14	ZooKeeper Befehlszeilen-Client.

## 6.3.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Aktionen zur Neukonfiguration treten auf, wenn Sie eine Konfiguration für Instance-Gruppen in einem laufenden Cluster angeben. Amazon EMR initiiert nur Rekonfigurationsaktionen für die Klassifizierungen, die Sie ändern. Weitere Informationen finden Sie unter [Eine Instance-Gruppe in einem laufenden Cluster neu konfigurieren](#).

### emr-6.3.0-Klassifizierungen

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.	Restarts the Resource Manager service.
container-executor	Ändern Sie die Werte in der Datei „ <code>container-executor.cfg</code> “ Datei von Hadoop YARN.	Not available.
container-log4j	Ändert die Werte in der <code>container-log4j.properties</code> -Datei in Hadoop YARN.	Not available.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
		Httpfs, and MapReduce-HistoryServer.
docker-conf	Ändern Sie die Docker-bezogenen Einstellungen.	Not available.
emrfs-site	Ändert die EMRFS-Einstellungen.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegistrator, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.	Restarts Flink history server.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.	Restarts Flink history server.
flink-log4j-session	Ändern Sie die Einstellungen von Flink log4j-session.properties für Kubernetes-/Yarn-Sitzungen.	Restarts Flink history server.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.	Restarts Flink history server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.	Not available.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Ändert die Werte in der HBase-Umgebung.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.	Not available.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.	This classification should not be reconfigured.
hdfs-env	Ändert die Werte in der HDFS-Umgebung.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei proto-hive-site .xml von HCatalog.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.	Restarts Hive WebHCat server.
hive	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Hive.	Sets configurations to launch Hive LLAP service.



Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.	Not available.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.	Not available.
hive-env	Ändert die Werte in der Hive-Umgebung.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.	Not available.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log	Not available.
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.	Not available.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hue-ini	Ändert die Werte in der INI-Datei in Hue.	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.	Restarts Hadoop Httpfs service.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.	Not available.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.	Not available.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Ändern der Werte in der Hudi-Umgebung.	Not available.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.	Not available.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.	Not available.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.	Not available.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.	Restarts Livy Server.
livy-env	Ändert die Werte in der Livy-Umgebung.	Restarts Livy Server.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.	Restarts Livy Server.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Ändert die Werte in der Oozie-Umgebung.	Restarts Oozie.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.	Restarts Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.	Restarts Oozie.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
phoenix-hbase-metrics	Ändert die Werte in der <code>hadoop-metrics2-hbase.properties</code> -Datei in Phoenix.	Not available.
phoenix-hbase-site	Ändert die Werte in der <code>hbase-site.xml</code> -Datei in Phoenix.	Not available.
phoenix-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Ändert die Werte in der <code>hadoop-metrics2-phoenix.properties</code> -Datei in Phoenix.	Not available.
pig-env	Ändert die Werte in der Pig-Umgebung.	Not available.
pig-properties	Ändert die Werte in der <code>pig.properties</code> -Datei in Pig.	Restarts Oozie.
pig-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Pig.	Not available.
presto-log	Ändert die Werte in der <code>log.properties</code> -Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Ändert die Werte in der <code>config.properties</code> -Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei <code>password-authenticator.properties</code> .	Not available.
presto-env	Ändern Sie die Werte in der <code>presto-env.sh</code> -Datei in Presto.	Restarts Presto-Server (for PrestoDB)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.	Not available.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.	Not available.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.	Not available.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.	Not available.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.	Not available.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.	Not available.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.	Not available.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.	Not available.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.	Not available.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.	Not available.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.	Not available.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.	Not available.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.	Not available.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.	Not available.
prestosql-log	Ändert die Werte in der log.properties-Datei in Presto.	Restarts Presto-Server (for PrestoSQL)
prestosql-config	Ändert die Werte in der config.properties-Datei in Presto.	Restarts Presto-Server (for PrestoSQL)
prestosql-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.	Restarts Presto-Server (for PrestoSQL)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
prestosql-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.	Restarts Presto-Server (for PrestoSQL)
prestosql-node	Ändern Sie die Werte in PrestoSQLs node.properties-Datei.	Not available.
prestosql-connector-blackhole	Ändern Sie die Werte in PrestoSQLs blackhole.properties-Datei.	Not available.
prestosql-connector-cassandra	Ändern Sie die Werte in PrestoSQLs cassandra.properties-Datei.	Not available.
prestosql-connector-hive	Ändern Sie die Werte in PrestoSQLs hive.properties-Datei.	Restarts Presto-Server (for PrestoSQL)
prestosql-connector-jmx	Ändern Sie die Werte in PrestoSQLs jmx.properties-Datei.	Not available.
prestosql-connector-kafka	Ändern Sie die Werte in PrestoSQLs kafka.properties-Datei.	Not available.
prestosql-connector-localfile	Ändern Sie die Werte in PrestoSQLs localfile.properties-Datei.	Not available.
prestosql-connector-memory	Ändern Sie die Werte in PrestoSQLs memory.properties-Datei in Presto.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
prestosql-connector-mongodb	Ändern Sie die Werte in PrestoSQLs mongod.properties-Datei in Presto.	Not available.
prestosql-connector-mysql	Ändern Sie die Werte in PrestoSQLs mysql.properties-Datei.	Not available.
prestosql-connector-postgresql	Ändern Sie die Werte in PrestoSQLs postgresql.properties-Datei.	Not available.
prestosql-connector-raptor	Ändern Sie die Werte in PrestoSQLs raptor.properties-Datei.	Not available.
prestosql-connector-redis	Ändern Sie die Werte in PrestoSQLs redis.properties-Datei.	Not available.
prestosql-connector-redshift	Ändern Sie die Werte in PrestoSQLs redshift.properties-Datei.	Not available.
prestosql-connector-tpch	Ändern Sie die Werte in PrestoSQLs tpch.properties-Datei.	Not available.
prestosql-connector-tpcds	Ändern Sie die Werte in PrestoSQLs tpcds.properties-Datei.	Not available.
ranger-kms-dbks-site	Ändert die Werte in der dbks-site.xml-Datei von Ranger KMS.	Restarts Ranger KMS Server.



Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Ändert die Werte in der Ranger KMS-Umgebung.	Restarts Ranger KMS Server.
ranger-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei von Ranger KMS.	Not available.
ranger-kms-db-ca	Ändert die Werte für die CA-Datei auf S3 für die MySQL SSL-Verbindung mit Ranger KMS.	Not available.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Ändert die Werte in der Spark-Umgebung.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.	Not available.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.	Not available.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.	Not available.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.	Not available.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.	Restart Oozie and HiveServer2.
yarn-env	Ändert die Werte in der YARN-Umgebung.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.	Restarts Zeppelin.
zeppelin-site	Ändern Sie die Konfigurationseinstellungen in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.	Restarts Zookeeper server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.	Restarts Zookeeper server.

## Amazon-EMR-Version 6.2.1

### 6.2.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [PrestoSQL](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

#### Informationen zur Anwendungsversion

	emr-6.2.1	emr-6.2.0	emr-6.1.1	emr-6.1.0
AWS SDK for Java	1.11.880	1.11.880	1.11.828	1.11.828
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2.12,10	2.12.10	2.12.10	2.12.10

	emr-6.2.1	emr-6.2.0	emr-6.1.1	emr-6.1.0
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.11.2	1.11.2	1.11.0	1.11.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.2.6-amzn-0	2.2.6-amzn-0	2.2.5	2.2.5
HCatalog	3.1.2	3.1.2	3.1.2	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.2	3.1.2	3.1.2	3.1.2
Hudi	0.6.0-amzn-1	0.6.0-amzn-1	0.5.2-incubating-amzn-2	0.5.2-incubating-amzn-2
Hue	4.8.0	4.8.0	4.7.1	4.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	2.1.0	2.1.0	-	-
JupyterHub	1.1.0	1.1.0	1.1.0	1.1.0
Livy	0.7.0	0.7.0	0.7.0	0.7.0
MXNet	1.7.0	1.7.0	1.6.0	1.6.0
Mahout	-	-	-	-
Oozie	5.2.0	5.2.0	5.2.0	5.2.0
Phoenix	5.0.0	5.0.0	5.0.0	5.0.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0

	emr-6.2.1	emr-6.2.0	emr-6.1.1	emr-6.1.0
Presto	0,238,3	0,238,3	0.232	0.232
Spark	3.0.1	3.0.1	3.0.0	3.0.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.3.1	2.3.1	2.1.0	2.1.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	343	343	338	338
Zeppelin	0.9.0	0.9.0	0.9.0	0.9.0
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

## 6.2.1 Versionshinweise

Dies ist eine Version zur Behebung von Problemen mit Amazon-EMR-Skalierung, wenn ein Cluster nicht erfolgreich hoch-/herunterskaliert werden kann oder Anwendungsfehler verursacht werden.

### Änderungen, Verbesserungen und behobene Probleme

- Es wurde ein Problem behoben, bei dem Skalierungsanforderungen für einen großen, stark ausgelasteten Cluster fehlschlagen, wenn Amazon-EMR-On-Cluster-Daemons Aktivitäten zur Integritätsprüfung durchführten, z. B. das Erfassen des YARN-Knotenstatus und des HDFS-Knotenstatus. Das lag daran, dass Cluster-Daemons die Integritätsstatusdaten eines Knotens nicht an interne Amazon-EMR-Komponenten weitergeben konnten.
- Verbesserte EMR-On-Cluster-Daemons zur korrekten Nachverfolgung der Knotenstatus bei der Wiederverwendung von IP-Adressen, um die Zuverlässigkeit bei Skalierungsvorgängen zu verbessern.
- [SPARK-29683](#). Es wurde ein Problem behoben, bei dem während der Cluster-Skalierung Auftragsfehler auftraten, da Spark davon ausging, dass alle verfügbaren Knoten auf der Verweigern-Liste standen.

- [YARN-9011](#). Es wurde ein Problem behoben, bei dem Auftragsfehler aufgrund eines Fehlers bei der Außerbetriebnahme von YARN auftraten, wenn der Cluster versuchte, hoch- oder herunterzuskalieren.
- Das Problem mit Schritt- oder Auftragsfehlern bei der Cluster-Skalierung wurde behoben, indem sichergestellt wurde, dass die Knotenstatus zwischen den Amazon-EMR-On-Cluster-Daemons und YARN/HDFS immer konsistent sind.
- Es wurde ein Problem behoben, bei dem Clustervorgänge wie Herunterskalierung und Schrittübermittlung für Amazon-EMR-Cluster, die mit Kerberos-Authentifizierung aktiviert waren, fehlschlugen. Dies lag daran, dass der Amazon-EMR-On-Cluster-Daemon das Kerberos-Ticket nicht erneuert hat, das für die sichere Kommunikation mit HDFS/YARN erforderlich ist, das auf dem Primärknoten ausgeführt wird.
- Neuere Amazon-EMR-Versionen beheben das Problem mit einem niedrigeren Limit für die „maximale Anzahl geöffneter Dateien“ auf älteren AL2-Versionen in Amazon EMR. Die Amazon-EMR-Versionen 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 und höher enthalten jetzt einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“.
- HTTPS ist jetzt standardmäßig für Amazon-Linux-Repositorys aktiviert. Wenn Sie eine Amazon-S3-VPCE-Richtlinie verwenden, um den Zugriff auf bestimmte Buckets zu beschränken, müssen Sie den neuen Amazon-Linux-Bucket-ARN `arn:aws:s3:::amazonlinux-2-repos-$region/*` zu Ihrer Richtlinie hinzufügen (ersetzen Sie `$region` durch die Region, in der sich der Endpunkt befindet). Weitere Informationen finden Sie unter diesem Thema in den AWS Diskussionsforen. [Ankündigung: Amazon Linux 2 unterstützt jetzt die Möglichkeit, HTTPS zu verwenden, wenn eine Verbindung zu Paket-Repositorys hergestellt wird.](#)

## Bekannte Probleme

- Wenn Sie Spark mit der Formatierung des Hive-Partitionsstandorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den Amazon-EMR-Versionen 5.30.0 bis 5.36.0 und 6.2.0 bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:
  - Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.
  - Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist `s3://bucket/table/p=a` ein Präfix von `s3://bucket/table/p=a b`.
  - Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen UTF-8-Wert, der kleiner als das `/`-Zeichen (U+002F) ist. Beispielsweise fällt das Leerzeichen (U+0020),

das in `s3://bucket/table/p=a b` zwischen `a` und `b` vorkommt, in diese Kategorie. Beachten Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: `!"#$%&'()*+,-.` Weitere Informationen finden Sie unter [UTF-8-Kodierungstabelle und Unicode-Zeichen](#).

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.

## 6.2.1 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.4.1	Amazon SageMaker Spark-SDK
<code>emr-ddb</code>	4.16.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	3.1.0	Praktische Bibliotheken für das Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-kinesis	3.5.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-notebook-env	1.0.0	Conda Env für EMR-Notebooks, das Jupyter Enterprise Gateway enthält
emr-s3-dist-cp	2.16.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	2.0.0	EMR S3Select-Konnektor
emrfs	2,44,0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.11.2	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
flink-jobmanager-config	1.11.2	Verwaltung von Ressourcen auf EMR-Knoten für Apache JobManager Flink.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.



Komponente	Version	Beschreibung
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	3.2.1-amzn-2.1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	3.2.1-amzn-2.1	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	3.2.1-amzn-2.1	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	3.2.1-amzn-2.1	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-hdfs-journalnode	3.2.1-amzn-2.1	HDFS-Service zum Verwalten des Hadoop-Dateisystem journals auf HA-Clustern.
hadoop-httpfs-server	3.2.1-amzn-2.1	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	3.2.1-amzn-2.1	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	3.2.1-amzn-2.1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.

Komponente	Version	Beschreibung
hadoop-yarn-nodemanager	3.2.1-amzn-2.1	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	3.2.1-amzn-2.1	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	3.2.1-amzn-2.1	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	2.2.6-amzn-0	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	2.2.6-amzn-0	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	2.2.6-amzn-0	HBase-Befehlszeilen-Client.
hbase-rest-server	2.2.6-amzn-0	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	2.2.6-amzn-0	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	3.1.2-amzn-3	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.

Komponente	Version	Beschreibung
hcatalog-server	3.1.2-amzn-3	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	3.1.2-amzn-3	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	3.1.2-amzn-3	Hive-Befehlszeilen-Client.
hive-hbase	3.1.2-amzn-3	Hive-hbase client.
hive-metastore-server	3.1.2-amzn-3	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	3.1.2-amzn-3	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.6.0-amzn-1	Inkrementelles Verarbeitungs-Framework zur Stromversorgung der Datenpipeline bei geringer Latenz und hoher Effizienz.
hudi-presto	0.6.0-amzn-1	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hudi-prestosql	0.6.0-amzn-1	Bündel-Bibliothek zum Ausführen von PrestoSQL mit Hudi.

Komponente	Version	Beschreibung
hudi-spark	0.6.0-amzn-1	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.
hue-server	4.8.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	1.1.0	Multi-User-Server für Jupyter-Notebooks
livy-server	0.7.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP- und Reverse-Proxy-Server.
mxnet	1.7.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.64+	MariaDB-Datenbankserver.
nvidia-cuda	10.1.243	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.2.0	Oozie-Befehlszeilen-Client.
oozie-server	5.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	4.4.0	Open Source Computer Vision Library.
phoenix-library	5.0.0-HBase-2.0	Die Phoenix-Bibliotheken für den Server und den Client

Komponente	Version	Beschreibung
phoenix-query-server	5.0.0-HBase-2.0	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.238.3-amzn-1	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.238.3-amzn-1	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.238.3-amzn-1	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
prestoql-coordinator	343	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Prestoql-Worker.
prestoql-worker	343	Service für das Ausführen von Teilen einer Abfrage.
prestoql-client	343	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.

Komponente	Version	Beschreibung
r	3.4.3	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	2.0.0	Apache Ranger Key Management System
spark-client	3.0.1-amzn-0.1	Spark-Befehlszeilen-Clients.
spark-history-server	3.0.1-amzn-0.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	3.0.1-amzn-0.1	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	3.0.1-amzn-0.1	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
spark-rapids	0.2.0	Nvidia-Spark-RAPIDS-Plugin, das Apache Spark mit GPUs beschleunigt.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	2.3.1	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.2	Die Tez-YARN-Anwendung und -Bibliotheken.

Komponente	Version	Beschreibung
webserver	2.4.41+	Apache HTTP-Server.
zeppelin-server	0.9.0-Vorschau1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.14	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.14	ZooKeeper Befehlszeilen-Cli-ent.

## 6.2.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Aktionen zur Neukonfiguration treten auf, wenn Sie eine Konfiguration für Instance-Gruppen in einem laufenden Cluster angeben. Amazon EMR initiiert nur Rekonfigurationsaktionen für die Klassifizierungen, die Sie ändern. Weitere Informationen finden Sie unter [Eine Instance-Gruppe in einem laufenden Cluster neu konfigurieren](#).

### emr-6.2.1-Klassifizierungen

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.	Restarts the Resource Manager service.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
container-executor	Ändern Sie die Werte in der Datei „container-executor.cfg“ Datei von Hadoop YARN.	Not available.
container-log4j	Ändert die Werte in der container-log4j.properties-Datei in Hadoop YARN.	Not available.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Ändern Sie die Docker-bezogenen Einstellungen.	Not available.



Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
emrfs-site	Ändert die EMRFS-Einstellungen.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegistrator, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.	Not available.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.	Not available.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4j-yarn-session.properties.	Not available.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.	Not available.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Ändert die Werte in der HBase-Umgebung.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.	Not available.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.	This classification should not be reconfigured.
hdfs-env	Ändert die Werte in der HDFS-Umgebung.	Restarts Hadoop HDFS ZKFC.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.	Restarts Hive WebHCat server.
hive	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Hive.	Sets configurations to launch Hive LLAP service.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.	Not available.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.	Not available.
hive-env	Ändert die Werte in der Hive-Umgebung.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.	Not available.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log	Not available.
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.	Not available.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hue-ini	Ändert die Werte in der INI-Datei in Hue.	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.	Restarts Hadoop Httpfs service.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.	Not available.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.	Not available.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Ändern der Werte in der Hudi-Umgebung.	Not available.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.	Not available.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.	Not available.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.	Not available.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.	Restarts Livy Server.
livy-env	Ändert die Werte in der Livy-Umgebung.	Restarts Livy Server.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.	Restarts Livy Server.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Ändert die Werte in der Oozie-Umgebung.	Restarts Oozie.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.	Restarts Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.	Restarts Oozie.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
phoenix-hbase-metrics	Ändert die Werte in der <code>hadoop-metrics2-hbase.properties</code> -Datei in Phoenix.	Not available.
phoenix-hbase-site	Ändert die Werte in der <code>hbase-site.xml</code> -Datei in Phoenix.	Not available.
phoenix-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Ändert die Werte in der <code>hadoop-metrics2-phoenix.properties</code> -Datei in Phoenix.	Not available.
pig-env	Ändert die Werte in der Pig-Umgebung.	Not available.
pig-properties	Ändert die Werte in der <code>pig.properties</code> -Datei in Pig.	Restarts Oozie.
pig-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Pig.	Not available.
presto-log	Ändert die Werte in der <code>log.properties</code> -Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Ändert die Werte in der <code>config.properties</code> -Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei <code>password-authenticator.properties</code> .	Not available.
presto-env	Ändern Sie die Werte in der <code>presto-env.sh</code> -Datei in Presto.	Restarts Presto-Server (for PrestoDB)



Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.	Not available.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.	Not available.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.	Not available.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.	Not available.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.	Not available.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.	Not available.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.	Not available.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.	Not available.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.	Not available.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.	Not available.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.	Not available.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.	Not available.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.	Not available.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.	Not available.
prestosql-log	Ändert die Werte in der log.properties-Datei in Presto.	Restarts Presto-Server (for PrestoSQL)
prestosql-config	Ändert die Werte in der config.properties-Datei in Presto.	Restarts Presto-Server (for PrestoSQL)
prestosql-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.	Restarts Presto-Server (for PrestoSQL)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
prestosql-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.	Restarts Presto-Server (for PrestoSQL)
prestosql-node	Ändern Sie die Werte in PrestoSQLs node.properties-Datei.	Not available.
prestosql-connector-blackhole	Ändern Sie die Werte in PrestoSQLs blackhole.properties-Datei.	Not available.
prestosql-connector-cassandra	Ändern Sie die Werte in PrestoSQLs cassandra.properties-Datei.	Not available.
prestosql-connector-hive	Ändern Sie die Werte in PrestoSQLs hive.properties-Datei.	Restarts Presto-Server (for PrestoSQL)
prestosql-connector-jmx	Ändern Sie die Werte in PrestoSQLs jmx.properties-Datei.	Not available.
prestosql-connector-kafka	Ändern Sie die Werte in PrestoSQLs kafka.properties-Datei.	Not available.
prestosql-connector-localfile	Ändern Sie die Werte in PrestoSQLs localfile.properties-Datei.	Not available.
prestosql-connector-memory	Ändern Sie die Werte in PrestoSQLs memory.properties-Datei in Presto.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
prestosql-connector-mongodb	Ändern Sie die Werte in PrestoSQLs mongod.properties-Datei in Presto.	Not available.
prestosql-connector-mysql	Ändern Sie die Werte in PrestoSQLs mysql.properties-Datei.	Not available.
prestosql-connector-postgresql	Ändern Sie die Werte in PrestoSQLs postgresql.properties-Datei.	Not available.
prestosql-connector-raptor	Ändern Sie die Werte in PrestoSQLs raptor.properties-Datei.	Not available.
prestosql-connector-redis	Ändern Sie die Werte in PrestoSQLs redis.properties-Datei.	Not available.
prestosql-connector-redshift	Ändern Sie die Werte in PrestoSQLs redshift.properties-Datei.	Not available.
prestosql-connector-tpch	Ändern Sie die Werte in PrestoSQLs tpch.properties-Datei.	Not available.
prestosql-connector-tpcds	Ändern Sie die Werte in PrestoSQLs tpcds.properties-Datei.	Not available.
ranger-kms-dbks-site	Ändert die Werte in der dbks-site.xml-Datei von Ranger KMS.	Restarts Ranger KMS Server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Ändert die Werte in der Ranger KMS-Umgebung.	Restarts Ranger KMS Server.
ranger-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei von Ranger KMS.	Not available.
ranger-kms-db-ca	Ändert die Werte für die CA-Datei auf S3 für die MySQL SSL-Verbindung mit Ranger KMS.	Not available.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Ändert die Werte in der Spark-Umgebung.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.	Not available.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.	Not available.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.	Not available.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.	Not available.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.	Restart Oozie.
yarn-env	Ändert die Werte in der YARN-Umgebung.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.	Restarts Zeppelin.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.	Restarts Zookeeper server.

## Amazon-EMR-Version 6.2.0

### 6.2.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [PrestoSQL](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

#### Informationen zur Anwendungsversion

	emr-6.2.0	emr-6.1.1	emr-6.1.0	emr-6.0.1
AWS SDK for Java	1.11.880	1.11.828	1.11.828	1.11.711
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2.12,10	2.12.10	2.12.10	2.12.10
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.11.2	1.11.0	1.11.0	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2

	emr-6.2.0	emr-6.1.1	emr-6.1.0	emr-6.0.1
HBase	2.2.6-amzn-0	2.2.5	2.2.5	2.2.3
HCatalog	3.1.2	3.1.2	3.1.2	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.2	3.1.2	3.1.2	3.1.2
Hudi	0.6.0-amzn-1	0.5.2-incubating-amzn-2	0.5.2-incubating-amzn-2	0.5.0-incubating-amzn-1
Hue	4.8.0	4.7.1	4.7.1	4.4.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	2.1.0	-	-	-
JupyterHub	1.1.0	1.1.0	1.1.0	1.0.0
Livy	0.7.0	0.7.0	0.7.0	0.6.0
MXNet	1.7.0	1.6.0	1.6.0	1.5.1
Mahout	-	-	-	-
Oozie	5.2.0	5.2.0	5.2.0	5.1.0
Phoenix	5.0.0	5.0.0	5.0.0	5.0.0
Pig	0.17.0	0.17.0	0.17.0	-
Presto	0,238,3	0.232	0.232	0.230
Spark	3.0.1	3.0.0	3.0.0	2.4.4
Sqoop	1.4.7	1.4.7	1.4.7	-
TensorFlow	2.3.1	2.1.0	2.1.0	1.14.0



	emr-6.2.0	emr-6.1.1	emr-6.1.0	emr-6.0.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	343	338	338	-
Zeppelin	0.9.0	0.9.0	0.9.0	0.9.0
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

## 6.2.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 6.2.0. Änderungen beziehen sich auf Version 6.1.0.

Veröffentlichungsdatum: 9. Dezember 2020

Letzte Aktualisierung: 4. Oktober 2021

### Unterstützte Anwendungen

- AWS SDK for Java Ausführung 1.11.828
- emr-record-server Ausführung 1.7.0
- Flink-Version 1.11.2
- Ganglia Version 3.7.2
- Hadoop-Version 3.2.1-amzn-1
- HBase Version 2.2.6-amzn-0
- HB 1.0.0 ase-operator-tools
- HCatalog-Version 3.1.2-amzn-0
- Hive Version 3.1.2-amzn-3
- Hudi Version 0.6.0-amzn-1
- Hue Version 4.8.0
- JupyterHub Ausführung 1.1.0
- Livy-Version 0.7.0
- MXNet-Version 1.7.0

- Oozie-Version 5.2.0
- Phoenix Version 5.0.0
- Pig Version 0.17.0
- Presto Version 0.238.3-amzn-1
- PrestoSQL Version 343
- Spark Version 3.0.1-amzn-0
- Spark-Rapids 0.2.0
- TensorFlow Ausführung 2.3.1
- Zeppelin-Version 0.9.0-preview1
- Zookeeper Version 3.4.14
- Konnektoren und Treiber: DynamoDB Connector 4.16.0

## Neue Features

- HBase: Die Umbenennung in der Commit-Phase wurde entfernt und persistentes HFile-Tracking hinzugefügt. Weitere Informationen finden Sie unter [Persistentes HFile-Tracking](#) im Amazon-EMR-Versionshandbuch.
- HBase: Backported [Erstellen Sie eine Konfiguration, die das Zwischenspeichern von Blöcken bei der Komprimierung erzwingt.](#)
- PrestoDB: Verbesserungen beim dynamischen Partitionsbereinigen. Die regelbasierte Join Reorder funktioniert mit nicht partitionierten Daten.
- Verwaltete Richtlinien mit Geltungsbereich: Um den AWS bewährten Methoden zu entsprechen, hat Amazon EMR verwaltete Standardrichtlinien mit EMR-Geltungsbereich der Version v2 eingeführt, die als Ersatz für Richtlinien dienen, die nicht mehr unterstützt werden. Weitere Informationen finden Sie unter [Verwaltete Richtlinien von Amazon EMR](#).
- Unterstützungsstatus des Instance Metadata Service (IMDS) V2: Für Amazon EMR 6.2 oder höher verwenden Amazon-EMR-Komponenten IMDSv2 für alle IMDS-Aufrufe. Für IMDS-Aufrufe in Ihrem Anwendungscode können Sie sowohl IMDSv1 als auch IMDSv2 verwenden oder das IMDS so konfigurieren, dass es aus Sicherheitsgründen nur IMDSv2 verwendet. Wenn Sie IMDSv1 in früheren Amazon-EMR-6.x-Versionen deaktivieren, führt dies zu einem Cluster-Startup-Fehler.

## Änderungen, Verbesserungen und behobene Probleme

- Dies ist eine Version zur Behebung von Problemen mit Amazon-EMR-Skalierung, wenn ein Cluster nicht erfolgreich hoch-/herunterskaliert werden kann oder Anwendungsfehler verursacht werden.
- Es wurde ein Problem behoben, bei dem Skalierungsanforderungen für einen großen, stark ausgelasteten Cluster fehlschlagen, wenn Amazon-EMR-On-Cluster-Daemons Aktivitäten zur Integritätsprüfung durchführten, z. B. das Erfassen des YARN-Knotenstatus und des HDFS-Knotenstatus. Das lag daran, dass Cluster-Daemons die Integritätsstatusdaten eines Knotens nicht an interne Amazon-EMR-Komponenten weitergeben konnten.
- Verbesserte EMR-On-Cluster-Daemons zur korrekten Nachverfolgung der Knotenstatus bei der Wiederverwendung von IP-Adressen, um die Zuverlässigkeit bei Skalierungsvorgängen zu verbessern.
- [SPARK-29683](#). Es wurde ein Problem behoben, bei dem während der Cluster-Skalierung Auftragsfehler auftraten, da Spark davon ausging, dass alle verfügbaren Knoten auf der Verweigern-Liste standen.
- [YARN-9011](#). Es wurde ein Problem behoben, bei dem Auftragsfehler aufgrund eines Fehlers bei der Außerbetriebnahme von YARN auftraten, wenn der Cluster versuchte, hoch- oder herunterzuskalieren.
- Das Problem mit Schritt- oder Auftragsfehlern bei der Cluster-Skalierung wurde behoben, indem sichergestellt wurde, dass die Knotenstatus zwischen den Amazon-EMR-On-Cluster-Daemons und YARN/HDFS immer konsistent sind.
- Es wurde ein Problem behoben, bei dem Clustervorgänge wie Herunterskalierung und Schrittübermittlung für Amazon-EMR-Cluster, die mit Kerberos-Authentifizierung aktiviert waren, fehlschlagen. Dies lag daran, dass der Amazon-EMR-On-Cluster-Daemon das Kerberos-Ticket nicht erneuert hat, das für die sichere Kommunikation mit HDFS/YARN erforderlich ist, das auf dem Primärknoten ausgeführt wird.
- Neuere Amazon-EMR-Versionen beheben das Problem mit einem niedrigeren Limit für die „maximale Anzahl geöffneter Dateien“ auf älteren AL2-Versionen in Amazon EMR. Die Amazon-EMR-Versionen 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 und höher enthalten jetzt einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“.
- Spark: Leistungsverbesserungen in der Spark-Laufzeit.

## Bekannte Probleme

- Amazon EMR 6.2 hat falsche Berechtigungen für die Datei `libinstance-controller-java /etc/cron.d/` in EMR 6.2.0 festgelegt. Die Berechtigungen für die Datei sind `645 (-rw-r--r-x)`, obwohl sie `644 (-rw-r--r--)` sein sollten. Aus diesem Grund protokolliert Amazon-EMR-Version 6.2 keine Instance-State-Logs und das Verzeichnis `/emr/instance-logs` ist leer. Dieses Problem wurde in Amazon EMR 6.3.0 und höher behoben.

Zur Umgehung dieses Problems führen Sie das folgende Skript als Bootstrap-Aktion beim Clusterstart aus.

```
#!/bin/bash
sudo chmod 644 /etc/cron.d/libinstance-controller-java
```

- Für private Subnetz-Cluster von Amazon EMR 6.2.0 und 6.3.0 können Sie nicht auf die Ganglia-Webbenutzeroberfläche zugreifen. Sie erhalten die Fehlermeldung „Zugriff verweigert (403)“. Andere Web-Benutzeroberflächen wie Spark, Hue, Zeppelin JupyterHub, Livy und Tez funktionieren normal. Der Zugriff auf die Ganglia-Web-Benutzeroberfläche auf öffentlichen Subnetzclustern funktioniert ebenfalls normal. Um dieses Problem zu beheben, starten Sie den `httpd`-Service auf dem Primärknoten mit `sudo systemctl restart httpd` neu. Dieses Problem wurde in Amazon EMR 6.4.0 behoben.
- In Amazon EMR 6.2.0 gibt es ein Problem, bei dem `httpd` kontinuierlich ausfällt, wodurch Ganglia nicht verfügbar ist. Sie erhalten die Fehlermeldung „Es kann keine Verbindung zum Server hergestellt werden“. Um einen Cluster zu reparieren, der bereits mit diesem Problem läuft, stellen Sie eine SSH-Verbindung zum Cluster-Primärknoten her und fügen Sie die Zeile `Listen 80` zu der Datei `httpd.conf` hinzu, die sich unter `/etc/httpd/conf/httpd.conf` befindet. Dieses Problem wurde in Amazon EMR 6.3.0 behoben.
- `HTTPD` schlägt auf EMR-6.2.0-Clustern fehl, wenn Sie eine Sicherheitskonfiguration verwenden. Dadurch ist die Benutzeroberfläche der Ganglia-Webanwendung nicht verfügbar. Um auf die Benutzeroberfläche der Ganglia-Webanwendung zuzugreifen, fügen Sie `Listen 80` der `/etc/httpd/conf/httpd.conf`-Datei auf dem Primärknoten Ihres Clusters etwas hinzu. Informationen zum Herstellen einer Verbindung zu Ihrem Cluster finden Sie unter [Verbinden mit dem Primärknoten über SSH](#).

EMR Notebooks können auch keine Verbindung mit EMR-6.2.0-Clustern herstellen, wenn Sie eine Sicherheitskonfiguration verwenden. Das Notebook kann keine Kernel auflisten und Spark-

Aufträge nicht weiterleiten. Wir empfehlen, stattdessen EMR Notebooks mit einer anderen Version von Amazon EMR zu verwenden.

- Niedrigeres Limit für die „Max. Anzahl geöffneter Dateien“ auf älteren AL2-Versionen [in neueren Versionen behoben]. Amazon-EMR-Versionen: emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0 und emr-6.2.0 basieren auf älteren Versionen von Amazon Linux 2 (AL2), die eine niedrigere Ulimit-Einstellung für „Max. Anzahl geöffneter Dateien“ haben, wenn Amazon-EMR-Cluster mit dem Standard-AMI erstellt werden. Die Amazon-EMR-Versionen 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 und höher enthalten einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“. Versionen mit einem niedrigeren Limit für geöffnete Dateien verursachen beim Senden des Spark-Jobs den Fehler „Zu viele offene Dateien“. In den betroffenen Versionen hat das Amazon-EMR-Standard-AMI eine Ulimit-StandardEinstellung von 4 096 für „Max. Anzahl geöffneter Dateien“, was unter dem Dateilimit von 65 536 im neuesten Amazon-Linux-2-AMI liegt. Die niedrigere Ulimit-Einstellung für „Max. Anzahl geöffneter Dateien“ führt dazu, dass der Spark-Job fehlschlägt, wenn der Spark-Treiber und der Executor versuchen, mehr als 4 096 Dateien zu öffnen. Um das Problem zu beheben, verfügt Amazon EMR über ein Bootstrap Action (BA)-Skript, das die Ulimit-Einstellung bei der Cluster-Erstellung anpasst.

Wenn Sie eine ältere Amazon-EMR-Version verwenden, für die dieses Problem nicht dauerhaft behoben ist, können Sie mit der folgenden Problemumgehung das Instance-Controller-Ulimit explizit auf maximal 65 536 Dateien festlegen.

Explizit ein ulimit über die Befehlszeile setzen

1. Bearbeiten Sie `/etc/systemd/system/instance-controller.service`, um die folgenden Parameter zum Abschnitt Service hinzuzufügen.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Starten Sie neu InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Mithilfe der Bootstrap-Aktion (BA) ein Ulimit festlegen

Sie können auch ein Bootstrap-Aktionsskript (BA) verwenden, um das Ulimit für den Instance-Controller bei der Clustererstellung auf 65 536 Dateien zu konfigurieren.

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

• **⚠ Important**

Amazon EMR 6.1.0 und 6.2.0 beinhalten ein Leistungsproblem, das sich kritisch auf alle Hudi-Operationen zum Einfügen, Upsert und Löschen auswirken kann. Wenn Sie Hudi mit Amazon EMR 6.1.0 oder 6.2.0 verwenden möchten, sollten Sie sich an den AWS Support wenden, um ein gepatchtes Hudi RPM zu erhalten.

• **⚠ Important**

EMR-Cluster, auf denen Amazon-Linux- oder Amazon-Linux-2-AMIs (Amazon Machine Images) ausgeführt werden, verwenden das Standardverhalten von Amazon Linux und laden wichtige und kritische Kernel-Updates, die einen Neustart erfordern, nicht automatisch herunter und installieren sie. Dies ist dasselbe Verhalten wie bei anderen Amazon-EC2-Instances, die das standardmäßige Amazon-Linux-AMI ausführen. Wenn neue Amazon-Linux-Softwareupdates, die einen Neustart erfordern (wie Kernel-, NVIDIA- und CUDA-Updates), nach der Veröffentlichung einer Amazon-EMR-Version verfügbar werden, laden EMR-Cluster-Instances, die das Standard-AMI ausführen, diese Updates

nicht automatisch herunter und installieren sie. Um Kernel-Updates zu erhalten, können Sie [Ihr Amazon-EMR-AMI](#) so anpassen, dass es [das neueste Amazon-Linux-AMI verwendet](#).

- Maven-Artefakte von Amazon EMR 6.2.0 werden nicht veröffentlicht. Sie werden mit einer zukünftigen Version von Amazon EMR veröffentlicht.
- Die persistente HFile-Nachverfolgung mithilfe der HBase-Storefile-Systemtabelle unterstützt das Feature zur Replikation von HBase-Regionen nicht. Weitere Informationen zur HBase-Regionsreplikation finden Sie unter [Zeitleistenkonsistente, hochverfügbare Lesevorgänge](#).
- Unterschiede zwischen den Bucketing-Versionen von Amazon EMR 6.x und EMR 5.x Hive

EMR 5.x verwendet OOS Apache Hive 2, während in EMR 6.x OOS Apache Hive 3 verwendet wird. Die Open-Source-Version Hive2 verwendet Bucketing Version 1, während die Open-Source-Version Hive3 Bucketing Version 2 verwendet. Dieser Unterschied in der Bucketing-Version zwischen Hive 2 (EMR 5.x) und Hive 3 (EMR 6.x) bedeutet, dass Hive-Bucketing-Hashing anders funktioniert. Sehen Sie sich das folgende Beispiel an.

Die folgende Tabelle ist ein Beispiel, das in EMR 6.x bzw. EMR 5.x erstellt wurde.

```
-- Using following LOCATION in EMR 6.x
CREATE TABLE test_bucketing (id INT, desc STRING)
PARTITIONED BY (day STRING)
CLUSTERED BY(id) INTO 128 BUCKETS
LOCATION 's3://your-own-s3-bucket/emr-6-bucketing/';

-- Using following LOCATION in EMR 5.x
LOCATION 's3://your-own-s3-bucket/emr-5-bucketing/';
```

Dieselben Daten werden sowohl in EMR 6.x als auch in EMR 5.x eingefügt.

```
INSERT INTO test_bucketing PARTITION (day='01') VALUES(66, 'some_data');
INSERT INTO test_bucketing PARTITION (day='01') VALUES(200, 'some_data');
```

Die Überprüfung des S3-Speicherorts zeigt, dass der Name der Bucketing-Datei unterschiedlich ist, da sich die Hashing-Funktion zwischen EMR 6.x (Hive 3) und EMR 5.x (Hive 2) unterscheidet.

```
[hadoop@ip-10-0-0-122 ~]$ aws s3 ls s3://your-own-s3-bucket/emr-6-bucketing/day=01/
2020-10-21 20:35:16          13 000025_0
2020-10-21 20:35:22          14 000121_0
[hadoop@ip-10-0-0-122 ~]$ aws s3 ls s3://your-own-s3-bucket/emr-5-bucketing/day=01/
```

```
2020-10-21 20:32:07      13 000066_0
2020-10-21 20:32:51      14 000072_0
```

Sie können den Versionsunterschied auch erkennen, indem Sie den folgenden Befehl in der Hive-CLI in EMR 6.x ausführen. Beachten Sie, dass die Bucketing-Version 2 zurückgegeben wird.

```
hive> DESCRIBE FORMATTED test_bucketing;
...
Table Parameters:
  bucketing_version      2
...
```

- **Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung**

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in Amazon-EMR-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie dem Herunterskalieren oder der schrittweisen Übermittlung auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

Workaround:

- SSH als hadoop-Benutzer für den führenden Primärknoten des EMR-Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den hadoop-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

### Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen



konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

- Wenn Sie Spark mit der Formatierung des Hive-Partitionsstandorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den Amazon-EMR-Versionen 5.30.0 bis 5.36.0 und 6.2.0 bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:
  - Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.
  - Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist `s3://bucket/table/p=a` ein Präfix von `s3://bucket/table/p=a b`.
  - Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen UTF-8-Wert, der kleiner als das /-Zeichen (U+002F) ist. Beispielsweise fällt das Leerzeichen (U+0020), das in `s3://bucket/table/p=a b` zwischen a und b vorkommt, in diese Kategorie. Beachten Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: !"#%&'()\*+,-. Weitere Informationen finden Sie unter [UTF-8-Kodierungstabelle und Unicode-Zeichen](#).

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.

## 6.2.0 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1.4.1	Amazon SageMaker Spark-SDK
emr-ddb	4.16.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	3.1.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.5.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-notebook-env	1.0.0	Conda Env für EMR-Notebooks, das Jupyter Enterprise Gateway enthält
emr-s3-dist-cp	2.16.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	2.0.0	EMR S3Select-Konnektor
emrfs	2,44,0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.11.2	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
flink-jobmanager-config	1.11.2	Verwaltung von Ressourcen auf EMR-Knoten für Apache JobManager Flink.

Komponente	Version	Beschreibung
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	3.2.1-amzn-2	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	3.2.1-amzn-2	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	3.2.1-amzn-2	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	3.2.1-amzn-2	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-hdfs-journalnode	3.2.1-amzn-2	HDFS-Service zum Verwalten des Hadoop-Dateisystem journals auf HA-Clustern.

Komponente	Version	Beschreibung
hadoop-https-server	3.2.1-amzn-2	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	3.2.1-amzn-2	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	3.2.1-amzn-2	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	3.2.1-amzn-2	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resource-manager	3.2.1-amzn-2	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	3.2.1-amzn-2	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	2.2.6-amzn-0	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	2.2.6-amzn-0	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.

Komponente	Version	Beschreibung
hbase-client	2.2.6-amzn-0	HBase-Befehlszeilen-Client.
hbase-rest-server	2.2.6-amzn-0	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	2.2.6-amzn-0	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	3.1.2-amzn-3	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	3.1.2-amzn-3	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	3.1.2-amzn-3	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	3.1.2-amzn-3	Hive-Befehlszeilen-Client.
hive-hbase	3.1.2-amzn-3	Hive-hbase client.
hive-metastore-server	3.1.2-amzn-3	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	3.1.2-amzn-3	Service zur Annahme von Hive-Abfragen als Webanfragen.

Komponente	Version	Beschreibung
hudi	0.6.0-amzn-1	Inkrementelles Verarbeitungs-Framework zur Stromversorgung der Datenpipeline bei geringer Latenz und hoher Effizienz.
hudi-presto	0.6.0-amzn-1	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hudi-prestosql	0.6.0-amzn-1	Bündel-Bibliothek zum Ausführen von PrestoSQL mit Hudi.
hudi-spark	0.6.0-amzn-1	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.
hue-server	4.8.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	1.1.0	Multi-User-Server für Jupyter-Notebooks
livy-server	0.7.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP- und Reverse-Proxy-Server.
mxnet	1.7.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.64+	MariaDB-Datenbankserver.

Komponente	Version	Beschreibung
nvidia-cuda	10.1.243	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.2.0	Oozie-Befehlszeilen-Client.
oozie-server	5.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	4.4.0	Open Source Computer Vision Library.
phoenix-library	5.0.0-HBase-2.0	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	5.0.0-HBase-2.0	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.238.3-amzn-1	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.238.3-amzn-1	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.238.3-amzn-1	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.

Komponente	Version	Beschreibung
prestosql-coordinator	343	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Prestosql-Worker.
prestosql-worker	343	Service für das Ausführen von Teilen einer Abfrage.
prestosql-client	343	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.3	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	2.0.0	Apache Ranger Key Management System
spark-client	3.0.1-amzn-0	Spark-Befehlszeilen-Clients.
spark-history-server	3.0.1-amzn-0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	3.0.1-amzn-0	In-Memory-Ausführungs-Engine für YARN.



Komponente	Version	Beschreibung
spark-yarn-slave	3.0.1-amzn-0	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
spark-rapids	0.2.0	Nvidia-Spark-RAPIDS-Plugin, das Apache Spark mit GPUs beschleunigt.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	2.3.1	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.2	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.41+	Apache HTTP-Server.
zeppelin-server	0.9.0-Vorschau1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.14	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.14	ZooKeeper Befehlszeilen-Client.

## 6.2.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Aktionen zur Neukonfiguration treten auf, wenn Sie eine Konfiguration für Instance-Gruppen in einem laufenden Cluster angeben. Amazon EMR initiiert nur Rekonfigurationsaktionen für die Klassifizierungen, die Sie ändern. Weitere Informationen finden Sie unter [Eine Instance-Gruppe in einem laufenden Cluster neu konfigurieren](#).

### emr-6.2.0-Klassifizierungen

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.	Restarts the Resource Manager service.
container-executor	Ändern Sie die Werte in der Datei „ <code>container-executor.cfg</code> “ Datei von Hadoop YARN.	Not available.
container-log4j	Ändert die Werte in der <code>container-log4j.properties</code> -Datei in Hadoop YARN.	Not available.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
		Httpfs, and MapReduce-HistoryServer.
docker-conf	Ändern Sie die Docker-bezogenen Einstellungen.	Not available.
emrfs-site	Ändert die EMRFS-Einstellungen.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegistratorserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.	Not available.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.	Not available.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4j-yarn-session.properties.	Not available.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.	Not available.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Ändert die Werte in der HBase-Umgebung.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.	Not available.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.	This classification should not be reconfigured.
hdfs-env	Ändert die Werte in der HDFS-Umgebung.	Restarts Hadoop HDFS ZKFC.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.	Restarts Hive WebHCat server.
hive	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Hive.	Sets configurations to launch Hive LLAP service.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.	Not available.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.	Not available.
hive-env	Ändert die Werte in der Hive-Umgebung.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.	Not available.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log	Not available.
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.	Not available.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hue-ini	Ändert die Werte in der INI-Datei in Hue.	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.	Restarts Hadoop Httpfs service.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.	Not available.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.	Not available.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Ändern der Werte in der Hudi-Umgebung.	Not available.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.	Not available.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.	Not available.



Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.	Not available.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.	Not available.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.	Restarts Livy Server.
livy-env	Ändert die Werte in der Livy-Umgebung.	Restarts Livy Server.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.	Restarts Livy Server.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Ändert die Werte in der Oozie-Umgebung.	Restarts Oozie.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.	Restarts Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.	Restarts Oozie.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
phoenix-hbase-metrics	Ändert die Werte in der <code>hadoop-metrics2-hbase.properties</code> -Datei in Phoenix.	Not available.
phoenix-hbase-site	Ändert die Werte in der <code>hbase-site.xml</code> -Datei in Phoenix.	Not available.
phoenix-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Ändert die Werte in der <code>hadoop-metrics2-phoenix.properties</code> -Datei in Phoenix.	Not available.
pig-env	Ändert die Werte in der Pig-Umgebung.	Not available.
pig-properties	Ändert die Werte in der <code>pig.properties</code> -Datei in Pig.	Restarts Oozie.
pig-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Pig.	Not available.
presto-log	Ändert die Werte in der <code>log.properties</code> -Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Ändert die Werte in der <code>config.properties</code> -Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei <code>password-authenticator.properties</code> .	Not available.
presto-env	Ändern Sie die Werte in der <code>presto-env.sh</code> -Datei in Presto.	Restarts Presto-Server (for PrestoDB)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.	Not available.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.	Not available.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.	Not available.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.	Not available.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.	Not available.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.	Not available.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.	Not available.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.	Not available.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.	Not available.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.	Not available.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.	Not available.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.	Not available.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.	Not available.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.	Not available.
prestosql-log	Ändert die Werte in der log.properties-Datei in Presto.	Restarts Presto-Server (for PrestoSQL)
prestosql-config	Ändert die Werte in der config.properties-Datei in Presto.	Restarts Presto-Server (for PrestoSQL)
prestosql-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.	Restarts Presto-Server (for PrestoSQL)

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
prestosql-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.	Restarts Presto-Server (for PrestoSQL)
prestosql-node	Ändern Sie die Werte in PrestoSQLs node.properties-Datei.	Not available.
prestosql-connector-blackhole	Ändern Sie die Werte in PrestoSQLs blackhole.properties-Datei.	Not available.
prestosql-connector-cassandra	Ändern Sie die Werte in PrestoSQLs cassandra.properties-Datei.	Not available.
prestosql-connector-hive	Ändern Sie die Werte in PrestoSQLs hive.properties-Datei.	Restarts Presto-Server (for PrestoSQL)
prestosql-connector-jmx	Ändern Sie die Werte in PrestoSQLs jmx.properties-Datei.	Not available.
prestosql-connector-kafka	Ändern Sie die Werte in PrestoSQLs kafka.properties-Datei.	Not available.
prestosql-connector-localfile	Ändern Sie die Werte in PrestoSQLs localfile.properties-Datei.	Not available.
prestosql-connector-memory	Ändern Sie die Werte in PrestoSQLs memory.properties-Datei in Presto.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
prestosql-connector-mongodb	Ändern Sie die Werte in PrestoSQLs mongod.properties-Datei in Presto.	Not available.
prestosql-connector-mysql	Ändern Sie die Werte in PrestoSQLs mysql.properties-Datei.	Not available.
prestosql-connector-postgresql	Ändern Sie die Werte in PrestoSQLs postgresql.properties-Datei.	Not available.
prestosql-connector-raptor	Ändern Sie die Werte in PrestoSQLs raptor.properties-Datei.	Not available.
prestosql-connector-redis	Ändern Sie die Werte in PrestoSQLs redis.properties-Datei.	Not available.
prestosql-connector-redshift	Ändern Sie die Werte in PrestoSQLs redshift.properties-Datei.	Not available.
prestosql-connector-tpch	Ändern Sie die Werte in PrestoSQLs tpch.properties-Datei.	Not available.
prestosql-connector-tpcds	Ändern Sie die Werte in PrestoSQLs tpcds.properties-Datei.	Not available.
ranger-kms-dbks-site	Ändert die Werte in der dbks-site.xml-Datei von Ranger KMS.	Restarts Ranger KMS Server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Ändert die Werte in der Ranger KMS-Umgebung.	Restarts Ranger KMS Server.
ranger-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei von Ranger KMS.	Not available.
ranger-kms-db-ca	Ändert die Werte für die CA-Datei auf S3 für die MySQL SSL-Verbindung mit Ranger KMS.	Not available.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Ändert die Werte in der Spark-Umgebung.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.	Not available.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.	Not available.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.	Not available.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.	Not available.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.	Restart Oozie.
yarn-env	Ändert die Werte in der YARN-Umgebung.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.	Restarts Zeppelin.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.	Restarts Zookeeper server.



## Amazon-EMR-Version 6.1.1

### 6.1.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [PrestoSQL](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

#### Informationen zur Anwendungsversion

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
AWS SDK for Java	1.11.828	1.11.828	1.11.711	1.11.711
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2.12,10	2.12.10	2.12.10	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.11.0	1.11.0	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
HBase	2.2.5	2.2.5	2.2.3	2.2.3
HCatalog	3.1.2	3.1.2	3.1.2	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.2	3.1.2	3.1.2	3.1.2
Hudi	0.5.2-incubating-amzn-2	0.5.2-incubating-amzn-2	0.5.0-incubating-amzn-1	0.5.0-incubating-amzn-1
Hue	4.7.1	4.7.1	4.4.0	4.4.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.1.0	1.1.0	1.0.0	1.0.0
Livy	0.7.0	0.7.0	0.6.0	0.6.0
MXNet	1.6.0	1.6.0	1.5.1	1.5.1
Mahout	-	-	-	-
Oozie	5.2.0	5.2.0	5.1.0	5.1.0
Phoenix	5.0.0	5.0.0	5.0.0	5.0.0
Pig	0.17.0	0.17.0	-	-
Presto	0.232	0.232	0.230	0.230
Spark	3.0.0	3.0.0	2.4.4	2.4.4
Sqoop	1.4.7	1.4.7	-	-
TensorFlow	2.1.0	2.1.0	1.14.0	1.14.0

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	338	338	-	-
Zeppelin	0.9.0	0.9.0	0.9.0	0.9.0
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

### 6.1.1 Versionshinweise

Dies ist eine Version zur Behebung von Problemen mit Amazon-EMR-Skalierung, wenn ein Cluster nicht erfolgreich hoch-/herunterskaliert werden kann oder Anwendungsfehler verursacht werden.

#### Änderungen, Verbesserungen und behobene Probleme

- Es wurde ein Problem behoben, bei dem Skalierungsanforderungen für einen großen, stark ausgelasteten Cluster fehlschlugen, wenn Amazon-EMR-On-Cluster-Daemons Aktivitäten zur Integritätsprüfung durchführten, z. B. das Erfassen des YARN-Knotenstatus und des HDFS-Knotenstatus. Das lag daran, dass Cluster-Daemons die Integritätsstatusdaten eines Knotens nicht an interne Amazon-EMR-Komponenten weitergeben konnten.
- Verbesserte EMR-On-Cluster-Daemons zur korrekten Nachverfolgung der Knotenstatus bei der Wiederverwendung von IP-Adressen, um die Zuverlässigkeit bei Skalierungsvorgängen zu verbessern.
- [SPARK-29683](#). Es wurde ein Problem behoben, bei dem während der Cluster-Skalierung Auftragsfehler auftraten, da Spark davon ausging, dass alle verfügbaren Knoten auf der Verweigern-Liste standen.
- [YARN-9011](#). Es wurde ein Problem behoben, bei dem Auftragsfehler aufgrund eines Fehlers bei der Außerbetriebnahme von YARN auftraten, wenn der Cluster versuchte, hoch- oder herunterzuskalieren.
- Das Problem mit Schritt- oder Auftragsfehlern bei der Cluster-Skalierung wurde behoben, indem sichergestellt wurde, dass die Knotenstatus zwischen den Amazon-EMR-On-Cluster-Daemons und YARN/HDFS immer konsistent sind.
- Es wurde ein Problem behoben, bei dem Clustervorgänge wie Herunterskalierung und Schrittübermittlung für Amazon-EMR-Cluster, die mit Kerberos-Authentifizierung aktiviert waren,

fehlschlugen. Dies lag daran, dass der Amazon-EMR-On-Cluster-Daemon das Kerberos-Ticket nicht erneuert hat, das für die sichere Kommunikation mit HDFS/YARN erforderlich ist, das auf dem Primärknoten ausgeführt wird.

- Neuere Amazon-EMR-Versionen beheben das Problem mit einem niedrigeren Limit für die „maximale Anzahl geöffneter Dateien“ auf älteren AL2-Versionen in Amazon EMR. Die Amazon-EMR-Versionen 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 und höher enthalten jetzt einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“.
- HTTPS ist jetzt standardmäßig für Amazon-Linux-Repositorys aktiviert. Wenn Sie eine Amazon-S3-VPCE-Richtlinie verwenden, um den Zugriff auf bestimmte Buckets zu beschränken, müssen Sie den neuen Amazon-Linux-Bucket-ARN `arn:aws:s3:::amazonlinux-2-repos-$region/*` zu Ihrer Richtlinie hinzufügen (ersetzen Sie `$region` durch die Region, in der sich der Endpunkt befindet). Weitere Informationen finden Sie unter diesem Thema in den AWS Diskussionsforen. [Ankündigung: Amazon Linux 2 unterstützt jetzt die Möglichkeit, HTTPS zu verwenden, wenn eine Verbindung zu Paket-Repositorys hergestellt wird.](#)

### 6.1.1 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.3.0	Amazon SageMaker Spark-SDK

Komponente	Version	Beschreibung
emr-ddb	4.14.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	3.1.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.5.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.14.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	2.0.0	EMR S3Select-Konnektor
emrfs	2.42,0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.11.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.

Komponente	Version	Beschreibung
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	3.2.1-amzn-1.1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	3.2.1-amzn-1.1	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	3.2.1-amzn-1.1	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	3.2.1-amzn-1.1	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-hdfs-journalnode	3.2.1-amzn-1.1	HDFS-Service zum Verwalten des Hadoop-Dateisystem journals auf HA-Clustern.
hadoop-httpfs-server	3.2.1-amzn-1.1	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	3.2.1-amzn-1.1	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	3.2.1-amzn-1.1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.

Komponente	Version	Beschreibung
hadoop-yarn-nodemanager	3.2.1-amzn-1.1	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	3.2.1-amzn-1.1	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	3.2.1-amzn-1.1	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	2.2.5	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	2.2.5	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	2.2.5	HBase-Befehlszeilen-Client.
hbase-rest-server	2.2.5	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	2.2.5	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	3.1.2-amzn-2	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.

Komponente	Version	Beschreibung
hcatalog-server	3.1.2-amzn-2	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	3.1.2-amzn-2	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	3.1.2-amzn-2	Hive-Befehlszeilen-Client.
hive-hbase	3.1.2-amzn-2	Hive-hbase client.
hive-metastore-server	3.1.2-amzn-2	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	3.1.2-amzn-2	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.5.2-incubating-amzn-2	Inkrementelles Verarbeitungs-Framework zur Stromversorgung der Datenpipeline bei geringer Latenz und hoher Effizienz.
hudi-presto	0.5.2-incubating-amzn-2	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hudi-prestosql	0.5.2-incubating-amzn-2	Bündel-Bibliothek zum Ausführen von PrestoSQL mit Hudi.



Komponente	Version	Beschreibung
hudi-spark	0.5.2-incubating-amzn-2	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.
hue-server	4.7.1	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	1.1.0	Multi-User-Server für Jupyter-Notebooks
livy-server	0.7.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP- und Reverse-Proxy-Server.
mxnet	1.6.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.64+	MariaDB-Datenbankserver.
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.2.0	Oozie-Befehlszeilen-Client.
oozie-server	5.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	4.3.0	Open Source Computer Vision Library.
phoenix-library	5.0.0-HBase-2.0	Die Phoenix-Bibliotheken für den Server und den Client

Komponente	Version	Beschreibung
phoenix-query-server	5.0.0-HBase-2.0	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.232	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.232	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.232	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
prestoql-coordinator	338	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Prestoql-Worker.
prestoql-worker	338	Service für das Ausführen von Teilen einer Abfrage.
prestoql-client	338	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.

Komponente	Version	Beschreibung
r	3.4.3	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	2.0.0	Apache Ranger Key Management System
spark-client	3.0.0-amzn-0.1	Spark-Befehlszeilen-Clients.
spark-history-server	3.0.0-amzn-0.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	3.0.0-amzn-0.1	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	3.0.0-amzn-0.1	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	2.1.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.2	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.41+	Apache HTTP-Server.

Komponente	Version	Beschreibung
zeppelin-server	0.9.0-Vorschau1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.14	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.14	ZooKeeper Befehlszeilen-Cli ent.

### 6.1.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

#### emr-6.1.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-executor	Ändern Sie die Werte in der Datei „ <code>container-executor.cfg</code> “ Datei von Hadoop YARN.
container-log4j	Ändert die Werte in der <code>container-log4j.properties</code> -Datei in Hadoop YARN.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.

Klassifizierungen	Beschreibung
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4 j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.

Klassifizierungen	Beschreibung
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-env	Ändert die Werte in der HDFS-Umgebung.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Hive.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.

Klassifizierungen	Beschreibung
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
hudi-env	Ändern der Werte in der Hudi-Umgebung.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.

Klassifizierungen	Beschreibung
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei <code>jupyterhub_config.py</code> .
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der <code>config.json</code> -Datei in Sparkmagic.
livy-conf	Ändert die Werte in der <code>livy.conf</code> -Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei <code>mapred-site.xml</code> der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der <code>oozie-log4j.properties</code> -Datei in Oozie.
oozie-site	Ändert die Werte in der <code>oozie-site.xml</code> -Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der <code>hadoop-metrics2-hbase.properties</code> -Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der <code>hbase-site.xml</code> -Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Phoenix.



Klassifizierungen	Beschreibung
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.
prestosql-log	Ändert die Werte in der log.properties-Datei in Presto.
prestosql-config	Ändert die Werte in der config.properties-Datei in Presto.

Klassifizierungen	Beschreibung
prestosql-password-authenticator	Ändern Sie Werte in der Presto-Datei <code>password-authenticator.properties</code> .
prestosql-env	Ändern Sie die Werte in der <code>prestosql-env.sh</code> -Datei in Presto.
prestosql-node	Ändern Sie die Werte in PrestoSQLs <code>node.properties</code> -Datei.
prestosql-connector-blackhole	Ändern Sie die Werte in PrestoSQLs <code>blackhole.properties</code> -Datei.
prestosql-connector-cassandra	Ändern Sie die Werte in PrestoSQLs <code>cassandra.properties</code> -Datei.
prestosql-connector-hive	Ändern Sie die Werte in PrestoSQLs <code>hive.properties</code> -Datei.
prestosql-connector-jmx	Ändern Sie die Werte in PrestoSQLs <code>jmx.properties</code> -Datei.
prestosql-connector-kafka	Ändern Sie die Werte in PrestoSQLs <code>kafka.properties</code> -Datei.
prestosql-connector-localfile	Ändern Sie die Werte in PrestoSQLs <code>localfile.properties</code> -Datei.
prestosql-connector-memory	Ändern Sie die Werte in PrestoSQLs <code>memory.properties</code> -Datei in Presto.
prestosql-connector-mongodb	Ändern Sie die Werte in PrestoSQLs <code>mongodb.properties</code> -Datei in Presto.
prestosql-connector-mysql	Ändern Sie die Werte in PrestoSQLs <code>mysql.properties</code> -Datei.
prestosql-connector-postgresql	Ändern Sie die Werte in PrestoSQLs <code>postgresql.properties</code> -Datei.

Klassifizierungen	Beschreibung
prestoql-connector-raptor	Ändern Sie die Werte in PrestoSQLs raptor.properties-Datei.
prestoql-connector-redis	Ändern Sie die Werte in PrestoSQLs redis.properties-Datei.
prestoql-connector-redshift	Ändern Sie die Werte in PrestoSQLs redshift.properties-Datei.
prestoql-connector-tpch	Ändern Sie die Werte in PrestoSQLs tpch.properties-Datei.
prestoql-connector-tpcds	Ändern Sie die Werte in PrestoSQLs tpcds.properties-Datei.
ranger-kms-dbks-site	Ändert die Werte in der dbks-site.xml-Datei von Ranger KMS.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger KMS.
ranger-kms-env	Ändert die Werte in der Ranger KMS-Umgebung.
ranger-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei von Ranger KMS.
ranger-kms-db-ca	Ändert die Werte für die CA-Datei auf S3 für die MySQL SSL-Verbindung mit Ranger KMS.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.

Klassifizierungen	Beschreibung
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 6.1.0

### 6.1.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [PrestoSQL](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

#### Informationen zur Anwendungsversion

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
AWS SDK for Java	1.11.828	1.11.828	1.11.711	1.11.711
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2.12,10	2.12.10	2.12.10	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.11.0	1.11.0	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
HBase	2.2.5	2.2.5	2.2.3	2.2.3
HCatalog	3.1.2	3.1.2	3.1.2	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.2	3.1.2	3.1.2	3.1.2
Hudi	0.5.2-incubating-amzn-2	0.5.2-incubating-amzn-2	0.5.0-incubating-amzn-1	0.5.0-incubating-amzn-1
Hue	4.7.1	4.7.1	4.4.0	4.4.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.1.0	1.1.0	1.0.0	1.0.0
Livy	0.7.0	0.7.0	0.6.0	0.6.0
MXNet	1.6.0	1.6.0	1.5.1	1.5.1
Mahout	-	-	-	-
Oozie	5.2.0	5.2.0	5.1.0	5.1.0
Phoenix	5.0.0	5.0.0	5.0.0	5.0.0
Pig	0.17.0	0.17.0	-	-
Presto	0.232	0.232	0.230	0.230
Spark	3.0.0	3.0.0	2.4.4	2.4.4
Sqoop	1.4.7	1.4.7	-	-
TensorFlow	2.1.0	2.1.0	1.14.0	1.14.0

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	338	338	-	-
Zeppelin	0.9.0	0.9.0	0.9.0	0.9.0
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

## 6.1.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 6.1.0. Änderungen beziehen sich auf Version 6.0.0.

Erste Version: 4. September 2020

Letzte Aktualisierung: 15. Oktober 2020

### Unterstützte Anwendungen

- AWS SDK for Java Ausführung 1.11.828
- Flink-Version 1.11.0
- Ganglia Version 3.7.2
- Hadoop-Version 3.2.1-amzn-1
- HBase-Version 2.2.5
- ase-operator-tools HB 1.0.0
- HCatalog-Version 3.1.2-amzn-0
- Hive-Version 3.1.2-amzn-1
- Hudi-Version 0.5.2-incubating
- Hue-Version 4.7.1
- JupyterHub Ausführung 1.1.0
- Livy-Version 0.7.0
- MXNet-Version 1.6.0
- Oozie-Version 5.2.0



- Phoenix Version 5.0.0
- Presto-Version 0.232
- PrestoSQL-Version 338
- Spark-Version 3.0.0-amzn-0
- TensorFlow Ausführung 2.1.0
- Zeppelin-Version 0.9.0-preview1
- Zookeeper Version 3.4.14
- Anschlüsse und Treiber: DynamoDB Connector 4.14.0

## Neue Features

- ARM-Instance-Typen werden ab Amazon-EMR-Version 5.30.0 und Amazon-EMR-Version 6.1.0 unterstützt.
- Allzweck-Instance-Typen von M6g werden ab den Amazon-EMR-Versionen 6.1.0 und 5.30.0 unterstützt. Weitere Informationen finden Sie unter [Instance-Flotten konfigurieren](#) im Verwaltungshandbuch für Amazon EMR.
- Das EC2-Platzierungsgruppenfeature wird ab Amazon-EMR-Version 5.23.0 als Option für Cluster mit mehreren Primärknoten unterstützt. Derzeit werden nur Primärknoten-Typen von des Platzierungsgruppenfeatures unterstützt, und die SPREAD-Strategie wird auf diese Primärknoten angewendet. Bei dieser SPREAD-Strategie wird eine kleine Gruppe von Instances auf separater zugrundeliegender Hardware platziert, um den Verlust mehrerer Primärknoten im Falle eines Hardwarefehlers zu verhindern. Weitere Informationen finden Sie unter [EMR-Integration mit EC2 Placement Group](#) im Verwaltungshandbuch für Amazon EMR.
- Verwaltete Skalierung – Mit Amazon EMR ab Version 6.1.0 können Sie Amazon EMR Managed Scaling aktivieren, um die Anzahl der Instances oder Einheiten in Ihrem Cluster basierend auf der Workload automatisch zu erhöhen oder zu verringern. Amazon EMR wertet Cluster-Metriken kontinuierlich aus, um Skalierungsentscheidungen zu treffen, die Ihre Cluster für Kosten und Geschwindigkeit optimieren. Verwaltete Skalierung ist auch in Amazon-EMR-Version 5.30.0 und höher verfügbar, außer 6.0.0. Weitere Informationen finden Sie unter [Skalieren von Clusterressourcen](#) im Verwaltungshandbuch für Amazon EMR.
- PrestoSQL-Version 338 wird mit EMR 6.1.0 unterstützt. Weitere Informationen finden Sie unter [Presto](#).
  - PrestoSQL wird nur auf EMR 6.1.0 und späteren Versionen unterstützt, nicht auf EMR 6.0.0 oder EMR 5.x.

- Der Anwendungsname `Presto` wird weiterhin verwendet, um PrestoDB auf Clustern zu installieren. Verwenden Sie den Anwendungsnamen `PrestoSQL`, um PrestoSQL auf Clustern zu installieren.
- Sie können entweder PrestoDB oder PrestoSQL installieren, aber Sie können nicht beide auf einem einzigen Cluster installieren. Wenn beim Versuch, einen Cluster zu erstellen, sowohl PrestoDB als auch PrestoSQL angegeben werden, tritt ein Validierungsfehler auf und die Clustererstellungsanforderung schlägt fehl.
- PrestoSQL wird sowohl auf Single-Haupt- als auch auf Multi-Haupt-Clustern unterstützt. Auf Multi-Haupt-Clustern ist ein externer Hive-Metastore erforderlich, um PrestoSQL oder PrestoDB auszuführen. Weitere Informationen finden Sie unter [Unterstützte Anwendungen in einem EMR-Cluster mit mehreren Primärknoten](#).
- Unterstützung der auto ECR-Authentifizierung auf Apache Hadoop und Apache Spark mit Docker: Spark-Benutzer können Docker-Images von Docker Hub und Amazon Elastic Container Registry (Amazon ECR) verwenden, um Umgebungs- und Bibliotheksabhängigkeiten zu definieren.

[Konfigurieren Sie Docker](#) und [führen Sie Spark-Anwendungen mit Docker mithilfe von Amazon EMR 6.x aus](#).

- EMR unterstützt Apache Hive ACID-Transaktionen: Amazon EMR 6.1.0 bietet Unterstützung für Hive-ACID-Transaktionen, sodass es den ACID-Eigenschaften einer Datenbank entspricht. Mit diesem Feature können Sie `INSERT`, `UPDATE`, `DELETE`, `-` und `MERGE`-Operationen in Hive-Tabellen mit Daten in Amazon Simple Storage Service (Amazon S3) ausführen. Dies ist ein wichtiges Feature für Anwendungsfälle wie Streaming-Erfassung, Neuformulierung von Daten, Massenaktualisierungen mit `MERGE` und sich langsam ändernde Dimensionen. Weitere Informationen, einschließlich Konfigurationsbeispielen und Anwendungsfällen, finden Sie unter [Amazon EMR unterstützt Apache-Hive-ACID-Transaktionen](#).

## Änderungen, Verbesserungen und behobene Probleme

- Dies ist eine Version zur Behebung von Problemen mit Amazon-EMR-Skalierung, wenn ein Cluster nicht erfolgreich hoch-/herunterskaliert werden kann oder Anwendungsfehler verursacht werden.
- Es wurde ein Problem behoben, bei dem Skalierungsanforderungen für einen großen, stark ausgelasteten Cluster fehlschlagen, wenn Amazon-EMR-On-Cluster-Daemons Aktivitäten zur Integritätsprüfung durchführten, z. B. das Erfassen des YARN-Knotenstatus und des HDFS-Knotenstatus. Das lag daran, dass Cluster-Daemons die Integritätsstatusdaten eines Knotens nicht an interne Amazon-EMR-Komponenten weitergeben konnten.

- Verbesserte EMR-On-Cluster-Daemons zur korrekten Nachverfolgung der Knotenstatus bei der Wiederverwendung von IP-Adressen, um die Zuverlässigkeit bei Skalierungsvorgängen zu verbessern.
- [SPARK-29683](#). Es wurde ein Problem behoben, bei dem während der Cluster-Skalierung Auftragsfehler auftraten, da Spark davon ausging, dass alle verfügbaren Knoten auf der Verweigern-Liste standen.
- [YARN-9011](#). Es wurde ein Problem behoben, bei dem Auftragsfehler aufgrund eines Fehlers bei der Außerbetriebnahme von YARN auftraten, wenn der Cluster versuchte, hoch- oder herunterzuskalieren.
- Das Problem mit Schritt- oder Auftragsfehlern bei der Cluster-Skalierung wurde behoben, indem sichergestellt wurde, dass die Knotenstatus zwischen den Amazon-EMR-On-Cluster-Daemons und YARN/HDFS immer konsistent sind.
- Es wurde ein Problem behoben, bei dem Clustervorgänge wie Herunterskalierung und Schrittübermittlung für Amazon-EMR-Cluster, die mit Kerberos-Authentifizierung aktiviert waren, fehlschlagen. Dies lag daran, dass der Amazon-EMR-On-Cluster-Daemon das Kerberos-Ticket nicht erneuert hat, das für die sichere Kommunikation mit HDFS/YARN erforderlich ist, das auf dem Primärknoten ausgeführt wird.
- Neuere Amazon-EMR-Versionen beheben das Problem mit einem niedrigeren Limit für die „maximale Anzahl geöffneter Dateien“ auf älteren AL2-Versionen in Amazon EMR. Die Amazon-EMR-Versionen 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 und höher enthalten jetzt einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“.
- Apache Flink wird auf EMR 6.0.0 nicht unterstützt, aber es wird auf EMR 6.1.0 mit Flink 1.11.0 unterstützt. Dies ist die erste Version von Flink, die Hadoop 3 offiziell unterstützt. Siehe Ankündigung der Veröffentlichung von [Apache Flink 1.11.0](#).
- Ganglia wurde aus den standardmäßigen EMR-6.1.0-Paketen entfernt.

## Bekannte Probleme

- Niedrigeres Limit für die „Max. Anzahl geöffneter Dateien“ auf älteren AL2-Versionen [in neueren Versionen behoben]. Amazon-EMR-Versionen: emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0 und emr-6.2.0 basieren auf älteren Versionen von Amazon Linux 2 (AL2), die eine niedrigere Ulimit-Einstellung für „Max. Anzahl geöffneter Dateien“ haben, wenn Amazon-EMR-Cluster mit dem Standard-AMI erstellt werden. Die Amazon-EMR-Versionen 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 und höher enthalten einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“. Versionen mit einem niedrigeren Limit für

geöffnete Dateien verursachen beim Senden des Spark-Jobs den Fehler „Zu viele offene Dateien“. In den betroffenen Versionen hat das Amazon-EMR-Standard-AMI eine Ulimit-StandardEinstellung von 4 096 für „Max. Anzahl geöffneter Dateien“, was unter dem Dateilimit von 65 536 im neuesten Amazon-Linux-2-AMI liegt. Die niedrigere Ulimit-Einstellung für „Max. Anzahl geöffneter Dateien“ führt dazu, dass der Spark-Job fehlschlägt, wenn der Spark-Treiber und der Executor versuchen, mehr als 4 096 Dateien zu öffnen. Um das Problem zu beheben, verfügt Amazon EMR über ein Bootstrap Action (BA)-Skript, das die Ulimit-Einstellung bei der Cluster-Erstellung anpasst.

Wenn Sie eine ältere Amazon-EMR-Version verwenden, für die dieses Problem nicht dauerhaft behoben ist, können Sie mit der folgenden Problemumgehung das Instance-Controller-Ulimit explizit auf maximal 65 536 Dateien festlegen.

Explizit ein ulimit über die Befehlszeile setzen

1. Bearbeiten Sie `/etc/systemd/system/instance-controller.service`, um die folgenden Parameter zum Abschnitt Service hinzuzufügen.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Starten Sie neu InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Mithilfe der Bootstrap-Aktion (BA) ein Ulimit festlegen

Sie können auch ein Bootstrap-Aktionsskript (BA) verwenden, um das Ulimit für den Instance-Controller bei der Clustererstellung auf 65 536 Dateien zu konfigurieren.

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
```

```
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

**⚠ Important**

Amazon EMR 6.1.0 und 6.2.0 beinhalten ein Leistungsproblem, das sich kritisch auf alle Hudi-Operationen zum Einfügen, Upsert und Löschen auswirken kann. Wenn Sie Hudi mit Amazon EMR 6.1.0 oder 6.2.0 verwenden möchten, sollten Sie sich an den AWS Support wenden, um ein gepatchtes Hudi RPM zu erhalten.

- Wenn Sie eine benutzerdefinierte Garbage-Collection-Konfiguration mit `spark.driver.extraJavaOptions` und `spark.executor.extraJavaOptions` festlegen, führt dies dazu, dass der Treiber/Executor-Start mit EMR 6.1 aufgrund einer widersprüchlichen Garbage-Collection-Konfiguration fehlschlägt. Mit EMR Version 6.1.0 sollten Sie stattdessen eine benutzerdefinierte Spark-Garbage-Collection-Konfiguration für Treiber und Executors mit den Eigenschaften `spark.driver.defaultJavaOptions` und `spark.executor.defaultJavaOptions` angeben. Weitere Informationen finden Sie unter [Apache Spark Runtime Environment](#) und [Konfigurieren von Spark Garbage Collection auf Amazon EMR 6.1.0](#).
- Die Verwendung von Pig mit Oozie (und innerhalb von Hue, da Hue Oozie-Aktionen zur Ausführung von Pig-Skripten verwendet) generiert den Fehler, dass eine native LZO-Bibliothek nicht geladen werden kann. Diese Fehlermeldung ist informativ und verhindert nicht, dass Pig ausgeführt wird.
- Hudi-Parallelitätsunterstützung: Derzeit unterstützt Hudi keine gleichzeitigen Schreibvorgänge in eine einzelne Hudi-Tabelle. Darüber hinaus macht Hudi alle Änderungen rückgängig, die von in Bearbeitung befindlichen Autoren vorgenommen wurden, bevor ein neuer Writer beginnen kann. Gleichzeitige Schreibvorgänge können diesen Mechanismus stören und zu Wettlaufbedingungen führen, was zu Datenbeschädigungen führen kann. Sie sollten sicherstellen, dass im Rahmen Ihres Datenverarbeitungs-Workflows immer nur ein einziger Hudi-Writer mit einer Hudi-Tabelle arbeitet. Hudi unterstützt mehrere gleichzeitige Lesegeräte, die mit derselben Hudi-Tabelle arbeiten.
- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in Amazon-EMR-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie dem Herunterskalieren oder der schrittweisen Übermittlung auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

Workaround:

- SSH als hadoop-Benutzer für den führenden Primärknoten des EMR-Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den hadoop-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

#### Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

- In Amazon EMR 6.1.0 gibt es ein Problem, das Cluster betrifft, auf denen Presto ausgeführt wird. Nach einem längeren Zeitraum (Tage) kann der Cluster Fehler wie „su: failed to execute /bin/bash: Resource temporarily unavailable“ oder „Shell-Anfrage auf Kanal 0 fehlgeschlagen“ ausgeben. Dieses Problem wird durch einen internen Amazon EMR-Prozess (InstanceController) verursacht, der zu viele Light Weight Processes (LWP) erzeugt, was letztendlich dazu führt, dass der Hadoop-Benutzer sein Nproc-Limit überschreitet. Dadurch wird verhindert, dass der Benutzer zusätzliche Prozesse öffnet. Die Lösung für dieses Problem ist ein Upgrade auf EMR 6.2.0.

## 6.1.0 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.3.0	Amazon SageMaker Spark-SDK
<code>emr-ddb</code>	4.14.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	3.1.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.5.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-s3-dist-cp</code>	2.14.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
<code>emr-s3-select</code>	2.0.0	EMR S3Select-Konnektor

Komponente	Version	Beschreibung
emrfs	2.42,0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.11.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	3.2.1-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	3.2.1-amzn-1	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	3.2.1-amzn-1	HDFS-Client und -Bibliothek für die Befehlszeile



Komponente	Version	Beschreibung
hadoop-hdfs-namenode	3.2.1-amzn-1	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-hdfs-journalnode	3.2.1-amzn-1	HDFS-Service zum Verwalten des Hadoop-Dateisystem journals auf HA-Clustern.
hadoop-httpfs-server	3.2.1-amzn-1	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	3.2.1-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	3.2.1-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	3.2.1-amzn-1	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	3.2.1-amzn-1	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	3.2.1-amzn-1	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.

Komponente	Version	Beschreibung
hbase-hmaster	2.2.5	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	2.2.5	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	2.2.5	HBase-Befehlszeilen-Client.
hbase-rest-server	2.2.5	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	2.2.5	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	3.1.2-amzn-2	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	3.1.2-amzn-2	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	3.1.2-amzn-2	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	3.1.2-amzn-2	Hive-Befehlszeilen-Client.
hive-hbase	3.1.2-amzn-2	Hive-hbase client.

Komponente	Version	Beschreibung
hive-metastore-server	3.1.2-amzn-2	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	3.1.2-amzn-2	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.5.2-incubating-amzn-2	Inkrementelles Verarbeitungs-Framework zur Stromversorgung der Datenpipeline bei geringer Latenz und hoher Effizienz.
hudi-presto	0.5.2-incubating-amzn-2	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hudi-prestosql	0.5.2-incubating-amzn-2	Bündel-Bibliothek zum Ausführen von PrestoSQL mit Hudi.
hudi-spark	0.5.2-incubating-amzn-2	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.
hue-server	4.7.1	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	1.1.0	Multi-User-Server für Jupyter-Notebooks
livy-server	0.7.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark

Komponente	Version	Beschreibung
nginx	1.12.1	nginx [engine x] ist ein HTTP- und Reverse-Proxy-Server.
mxnet	1.6.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.64+	MariaDB-Datenbankserver.
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.2.0	Oozie-Befehlszeilen-Client.
oozie-server	5.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	4.3.0	Open Source Computer Vision Library.
phoenix-library	5.0.0-HBase-2.0	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	5.0.0-HBase-2.0	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.232	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.232	Service für das Ausführen von Teilen einer Abfrage.

Komponente	Version	Beschreibung
presto-client	0.232	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
prestoql-coordinator	338	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Prestoql-Worker.
prestoql-worker	338	Service für das Ausführen von Teilen einer Abfrage.
prestoql-client	338	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.3	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	2.0.0	Apache Ranger Key Management System
spark-client	3.0.0-amzn-0	Spark-Befehlszeilen-Clients.
spark-history-server	3.0.0-amzn-0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.

Komponente	Version	Beschreibung
spark-on-yarn	3.0.0-amzn-0	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	3.0.0-amzn-0	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	2.1.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.2	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.41+	Apache HTTP-Server.
zeppelin-server	0.9.0-Vorschau1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.14	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.14	ZooKeeper Befehlszeilen-Client.

## 6.1.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

### emr-6.1.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-executor	Ändern Sie die Werte in der Datei „ <code>container-executor.cfg</code> “ Datei von Hadoop YARN.
container-log4j	Ändert die Werte in der <code>container-log4j.properties</code> -Datei in Hadoop YARN.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Ändert die <code>log4j-cli.properties</code> -Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Hadoop.

Klassifizierungen	Beschreibung
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-env	Ändert die Werte in der HDFS-Umgebung.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site



Klassifizierungen	Beschreibung
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Hive.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.

Klassifizierungen	Beschreibung
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
hudi-env	Ändern der Werte in der Hudi-Umgebung.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.

Klassifizierungen	Beschreibung
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.
prestosql-log	Ändert die Werte in der log.properties-Datei in Presto.
prestosql-config	Ändert die Werte in der config.properties-Datei in Presto.
prestosql-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
prestosql-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
prestosql-node	Ändern Sie die Werte in PrestoSQLs node.properties-Datei.
prestosql-connector-blackhole	Ändern Sie die Werte in PrestoSQLs blackhole.properties-Datei.
prestosql-connector-cassandra	Ändern Sie die Werte in PrestoSQLs cassandra.properties-Datei.

Klassifizierungen	Beschreibung
prestoql-connector-hive	Ändern Sie die Werte in PrestoSQLs hive.properties-Datei.
prestoql-connector-jmx	Ändern Sie die Werte in PrestoSQLs jmx.properties-Datei.
prestoql-connector-kafka	Ändern Sie die Werte in PrestoSQLs kafka.properties-Datei.
prestoql-connector-localfile	Ändern Sie die Werte in PrestoSQLs localfile.properties-Datei.
prestoql-connector-memory	Ändern Sie die Werte in PrestoSQLs memory.properties-Datei in Presto.
prestoql-connector-mongodb	Ändern Sie die Werte in PrestoSQLs mongod.properties-Datei in Presto.
prestoql-connector-mysql	Ändern Sie die Werte in PrestoSQLs mysql.properties-Datei.
prestoql-connector-postgresql	Ändern Sie die Werte in PrestoSQLs postgresql.properties-Datei.
prestoql-connector-raptor	Ändern Sie die Werte in PrestoSQLs raptor.properties-Datei.
prestoql-connector-redis	Ändern Sie die Werte in PrestoSQLs redis.properties-Datei.
prestoql-connector-redshift	Ändern Sie die Werte in PrestoSQLs redshift.properties-Datei.
prestoql-connector-tpch	Ändern Sie die Werte in PrestoSQLs tpch.properties-Datei.
prestoql-connector-tpcds	Ändern Sie die Werte in PrestoSQLs tpcds.properties-Datei.

Klassifizierungen	Beschreibung
ranger-kms-dbks-site	Ändert die Werte in der dbks-site.xml-Datei von Ranger KMS.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger KMS.
ranger-kms-env	Ändert die Werte in der Ranger KMS-Umgebung.
ranger-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei von Ranger KMS.
ranger-kms-db-ca	Ändert die Werte für die CA-Datei auf S3 für die MySQL SSL-Verbindung mit Ranger KMS.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.

Klassifizierungen	Beschreibung
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 6.0.1

### 6.0.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Presto](#), [Spark](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)



- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

## Informationen zur Anwendungsversion

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
AWS SDK for Java	1.11.828	1.11.828	1.11.711	1.11.711
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2.12,10	2.12.10	2.12.10	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.11.0	1.11.0	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.2.5	2.2.5	2.2.3	2.2.3
HCatalog	3.1.2	3.1.2	3.1.2	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.2	3.1.2	3.1.2	3.1.2
Hudi	0.5.2-incubating-amzn-2	0.5.2-incubating-amzn-2	0.5.0-incubating-amzn-1	0.5.0-incubating-amzn-1
Hue	4.7.1	4.7.1	4.4.0	4.4.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.1.0	1.1.0	1.0.0	1.0.0

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
Livy	0.7.0	0.7.0	0.6.0	0.6.0
MXNet	1.6.0	1.6.0	1.5.1	1.5.1
Mahout	-	-	-	-
Oozie	5.2.0	5.2.0	5.1.0	5.1.0
Phoenix	5.0.0	5.0.0	5.0.0	5.0.0
Pig	0.17.0	0.17.0	-	-
Presto	0.232	0.232	0.230	0.230
Spark	3.0.0	3.0.0	2.4.4	2.4.4
Sqoop	1.4.7	1.4.7	-	-
TensorFlow	2.1.0	2.1.0	1.14.0	1.14.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	338	338	-	-
Zeppelin	0.9.0	0.9.0	0.9.0	0.9.0
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

## 6.0.1 Versionshinweise

Dies ist eine Version zur Behebung von Problemen mit Amazon-EMR-Skalierung, wenn ein Cluster nicht erfolgreich hoch-/herunterskaliert werden kann oder Anwendungsfehler verursacht werden.

### Änderungen, Verbesserungen und behobene Probleme

- Es wurde ein Problem behoben, bei dem Skalierungsanforderungen für einen großen, stark ausgelasteten Cluster fehlschlagen, wenn Amazon-EMR-On-Cluster-Daemons Aktivitäten zur Integritätsprüfung durchführten, z. B. das Erfassen des YARN-Knotenstatus und des HDFS-

Knotenstatus. Das lag daran, dass Cluster-Daemons die Integritätsstatusdaten eines Knotens nicht an interne Amazon-EMR-Komponenten weitergeben konnten.

- Verbesserte EMR-On-Cluster-Daemons zur korrekten Nachverfolgung der Knotenstatus bei der Wiederverwendung von IP-Adressen, um die Zuverlässigkeit bei Skalierungsvorgängen zu verbessern.
- [SPARK-29683](#). Es wurde ein Problem behoben, bei dem während der Cluster-Skalierung Auftragsfehler auftraten, da Spark davon ausging, dass alle verfügbaren Knoten auf der Verweigern-Liste standen.
- [YARN-9011](#). Es wurde ein Problem behoben, bei dem Auftragsfehler aufgrund eines Fehlers bei der Außerbetriebnahme von YARN auftraten, wenn der Cluster versuchte, hoch- oder herunterzuskalieren.
- Das Problem mit Schritt- oder Auftragsfehlern bei der Cluster-Skalierung wurde behoben, indem sichergestellt wurde, dass die Knotenstatus zwischen den Amazon-EMR-On-Cluster-Daemons und YARN/HDFS immer konsistent sind.
- Es wurde ein Problem behoben, bei dem Clustervorgänge wie Herunterskalierung und Schrittübermittlung für Amazon-EMR-Cluster, die mit Kerberos-Authentifizierung aktiviert waren, fehlschlugen. Dies lag daran, dass der Amazon-EMR-On-Cluster-Daemon das Kerberos-Ticket nicht erneuert hat, das für die sichere Kommunikation mit HDFS/YARN erforderlich ist, das auf dem Primärknoten ausgeführt wird.
- Neuere Amazon-EMR-Versionen beheben das Problem mit einem niedrigeren Limit für die „maximale Anzahl geöffneter Dateien“ auf älteren AL2-Versionen in Amazon EMR. Die Amazon-EMR-Versionen 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 und höher enthalten jetzt einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“.
- HTTPS ist jetzt standardmäßig für Amazon-Linux-Repositorys aktiviert. Wenn Sie eine Amazon-S3-VPCE-Richtlinie verwenden, um den Zugriff auf bestimmte Buckets zu beschränken, müssen Sie den neuen Amazon-Linux-Bucket-ARN `arn:aws:s3:::amazonlinux-2-repos-$region/*` zu Ihrer Richtlinie hinzufügen (ersetzen Sie `$region` durch die Region, in der sich der Endpunkt befindet). Weitere Informationen finden Sie unter diesem Thema in den AWS Diskussionsforen. [Ankündigung: Amazon Linux 2 unterstützt jetzt die Möglichkeit, HTTPS zu verwenden, wenn eine Verbindung zu Paket-Repositorys hergestellt wird.](#)

## 6.0.1 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon

EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.2.6	Amazon SageMaker Spark-SDK
<code>emr-ddb</code>	4.14.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	3.0.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.5.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-s3-dist-cp</code>	2.14.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
<code>emr-s3-select</code>	1.5.0	EMR S3Select-Konnektor
<code>emrfs</code>	2.39.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	3.2.1-amzn-0.1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	3.2.1-amzn-0.1	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	3.2.1-amzn-0.1	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	3.2.1-amzn-0.1	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-hdfs-journalnode	3.2.1-amzn-0.1	HDFS-Service zum Verwalten des Hadoop-Dateisystem journals auf HA-Clustern.

Komponente	Version	Beschreibung
hadoop-httfs-server	3.2.1-amzn-0.1	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	3.2.1-amzn-0.1	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	3.2.1-amzn-0.1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	3.2.1-amzn-0.1	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	3.2.1-amzn-0.1	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	3.2.1-amzn-0.1	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	2.2.3	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	2.2.3	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.

Komponente	Version	Beschreibung
hbase-client	2.2.3	HBase-Befehlszeilen-Client.
hbase-rest-server	2.2.3	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	2.2.3	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	3.1.2-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	3.1.2-amzn-0	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	3.1.2-amzn-0	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	3.1.2-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	3.1.2-amzn-0	Hive-hbase client.
hive-metastore-server	3.1.2-amzn-0	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	3.1.2-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.

Komponente	Version	Beschreibung
hudi	0.5.0-incubating-amzn-1	Inkrementelles Verarbeitungs-Framework zur Stromversorgung der Datenpipeline bei geringer Latenz und hoher Effizienz.
hudi-presto	0.5.0-incubating-amzn-1	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hue-server	4.4.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	1.0.0	Multi-User-Server für Jupyter-Notebooks
livy-server	0.6.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP- und Reverse-Proxy-Server.
mxnet	1.5.1	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.64+	MariaDB-Datenbankserver.
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Toolkit
oozie-client	5.1.0	Oozie-Befehlszeilen-Client.
oozie-server	5.1.0	Service für die Annahme von Oozie Workflow-Anforderungen.



Komponente	Version	Beschreibung
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	5.0.0-HBase-2.0	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	5.0.0-HBase-2.0	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.230	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.230	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.230	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
r	3.4.3	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
spark-client	2.4.4	Spark-Befehlszeilen-Clients.

Komponente	Version	Beschreibung
spark-history-server	2.4.4	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.4.4	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.4.4	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
tensorflow	1.14.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.2	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.41+	Apache HTTP-Server.
zeppelin-server	0.9.0-SNAPSHOT	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.14	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.14	ZooKeeper Befehlszeilen-Client.

## 6.0.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

### emr-6.0.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-executor	Ändern Sie die Werte in der Datei „ <code>container-executor.cfg</code> “ Datei von Hadoop YARN.
container-log4j	Ändert die Werte in der <code>container-log4j.properties</code> -Datei in Hadoop YARN.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.

Klassifizierungen	Beschreibung
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-env	Ändert die Werte in der HDFS-Umgebung.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei proto-hive-site.xml von HCatalog.
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.

Klassifizierungen	Beschreibung
hive	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Hive.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.

Klassifizierungen	Beschreibung
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.

Klassifizierungen	Beschreibung
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.
ranger-kms-dbks-site	Ändert die Werte in der dbks-site.xml-Datei von Ranger KMS.



Klassifizierungen	Beschreibung
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger KMS.
ranger-kms-env	Ändert die Werte in der Ranger KMS-Umgebung.
ranger-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei von Ranger KMS.
ranger-kms-db-ca	Ändert die Werte für die CA-Datei auf S3 für die MySQL SSL-Verbindung mit Ranger KMS.
recordserver-env	Werte in der RecordServer EMR-Umgebung ändern.
recordserver-conf	Ändern Sie die Werte in der Datei RecordServer server.properties von EMR.
recordserver-log4j	Ändern Sie die Werte in der Datei RecordServer log4j.properties von EMR.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.

Klassifizierungen	Beschreibung
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 6.0.0

### 6.0.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Presto](#), [Spark](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

## Informationen zur Anwendungsversion

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
AWS SDK for Java	1.11.828	1.11.828	1.11.711	1.11.711
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2.12,10	2.12.10	2.12.10	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.11.0	1.11.0	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.2.5	2.2.5	2.2.3	2.2.3
HCatalog	3.1.2	3.1.2	3.1.2	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.2	3.1.2	3.1.2	3.1.2
Hudi	0.5.2-incubating-amzn-2	0.5.2-incubating-amzn-2	0.5.0-incubating-amzn-1	0.5.0-incubating-amzn-1
Hue	4.7.1	4.7.1	4.4.0	4.4.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.1.0	1.1.0	1.0.0	1.0.0
Livy	0.7.0	0.7.0	0.6.0	0.6.0

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
MXNet	1.6.0	1.6.0	1.5.1	1.5.1
Mahout	-	-	-	-
Oozie	5.2.0	5.2.0	5.1.0	5.1.0
Phoenix	5.0.0	5.0.0	5.0.0	5.0.0
Pig	0.17.0	0.17.0	-	-
Presto	0.232	0.232	0.230	0.230
Spark	3.0.0	3.0.0	2.4.4	2.4.4
Sqoop	1.4.7	1.4.7	-	-
TensorFlow	2.1.0	2.1.0	1.14.0	1.14.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	338	338	-	-
Zeppelin	0.9.0	0.9.0	0.9.0	0.9.0
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

## 6.0.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 6.0.0.

Erste Version: 10. März 2020

### Unterstützte Anwendungen

- AWS SDK for Java Ausführung 1.11.711
- Ganglia Version 3.7.2
- Hadoop Version 3.2.1

- HBase Version 2.2.3
- HCatalog Version 3.1.2
- Hive Version 3.1.2
- Hudi Version 0.5.0-incubating
- Hue Version 4.4.0
- JupyterHub Ausführung 1.0.0
- Livy Version 0.6.0
- MXNet-Version 1.5.1
- Oozie Version 5.1.0
- Phoenix Version 5.0.0
- Presto Version 0.230
- Spark Version 2.4.4
- TensorFlow Ausführung 1.14.0
- Zeppelin Version 0.9.0-SNAPSHOT
- Zookeeper Version 3.4.14
- Anschlüsse und Treiber: DynamoDB Connector 4.14.0

#### Note

Flink, Sqoop, Pig und Mahout sind in Amazon-EMR-Version 6.0.0 nicht verfügbar.

#### Neue Features

- YARN Docker Runtime Support - YARN-Anwendungen, wie etwa Spark-Aufgaben, können jetzt im Kontext eines Docker-Containers ausgeführt werden. Dadurch können Sie ganz einfach Abhängigkeiten in einem Docker-Image definieren, ohne dass benutzerdefinierte Bibliotheken auf dem Amazon-EMR-Cluster installiert werden müssen. Weitere Informationen finden Sie unter [Konfigurieren der Docker-Integration](#) und [Ausführen von Spark-Anwendungen mit Docker mit Amazon EMR 6.0.0](#).
- Unterstützung für Hive LLAP - Hive unterstützt jetzt den LLAP-Ausführungsmodus für eine verbesserte Abfrageleistung. Weitere Informationen finden Sie unter [Verwenden von Hive LLAP](#).

## Änderungen, Verbesserungen und behobene Probleme

- Dies ist eine Version zur Behebung von Problemen mit Amazon-EMR-Skalierung, wenn ein Cluster nicht erfolgreich hoch-/herunterskaliert werden kann oder Anwendungsfehler verursacht werden.
- Es wurde ein Problem behoben, bei dem Skalierungsanforderungen für einen großen, stark ausgelasteten Cluster fehlschlagen, wenn Amazon-EMR-On-Cluster-Daemons Aktivitäten zur Integritätsprüfung durchführten, z. B. das Erfassen des YARN-Knotenstatus und des HDFS-Knotenstatus. Das lag daran, dass Cluster-Daemons die Integritätsstatusdaten eines Knotens nicht an interne Amazon-EMR-Komponenten weitergeben konnten.
- Verbesserte EMR-On-Cluster-Daemons zur korrekten Nachverfolgung der Knotenstatus bei der Wiederverwendung von IP-Adressen, um die Zuverlässigkeit bei Skalierungsvorgängen zu verbessern.
- [SPARK-29683](#). Es wurde ein Problem behoben, bei dem während der Cluster-Skalierung Auftragsfehler auftraten, da Spark davon ausging, dass alle verfügbaren Knoten auf der Verweigern-Liste standen.
- [YARN-9011](#). Es wurde ein Problem behoben, bei dem Auftragsfehler aufgrund eines Fehlers bei der Außerbetriebnahme von YARN auftraten, wenn der Cluster versuchte, hoch- oder herunterzuskalieren.
- Das Problem mit Schritt- oder Auftragsfehlern bei der Cluster-Skalierung wurde behoben, indem sichergestellt wurde, dass die Knotenstatus zwischen den Amazon-EMR-On-Cluster-Daemons und YARN/HDFS immer konsistent sind.
- Es wurde ein Problem behoben, bei dem Clustervorgänge wie Herunterskalierung und Schrittübermittlung für Amazon-EMR-Cluster, die mit Kerberos-Authentifizierung aktiviert waren, fehlschlagen. Dies lag daran, dass der Amazon-EMR-On-Cluster-Daemon das Kerberos-Ticket nicht erneuert hat, das für die sichere Kommunikation mit HDFS/YARN erforderlich ist, das auf dem Primärknoten ausgeführt wird.
- Neuere Amazon-EMR-Versionen beheben das Problem mit einem niedrigeren Limit für die „maximale Anzahl geöffneter Dateien“ auf älteren AL2-Versionen in Amazon EMR. Die Amazon-EMR-Versionen 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 und höher enthalten jetzt einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“.
- Amazon Linux
  - Amazon Linux 2 ist das Betriebssystem für die EMR 6.x-Release-Serie.
  - `systemd` wird für die Serviceverwaltung statt `upstart` in Amazon Linux 1 verwendet.
- Java Development Kit (JDK)

- Coretto JDK 8 ist das Standard-JDK für die EMR-6.x-Release-Serie.
- Scala
  - Scala 2.12 wird mit Apache Spark und Apache Livy verwendet.
- Python 3
  - Python 3 ist jetzt die Standardversion von Python in EMR.
- YARN-Knotenbeschriftungen
  - Beginnend mit der Amazon-EMR-6.x-Release-Reihe ist das Feature YARN-Knotenbeschriftungen standardmäßig deaktiviert. Die Anwendungs-Master-Prozesse können standardmäßig sowohl auf Kern- als auch auf Aufgabenknoten ausgeführt werden. Sie können die Funktion YARN-Knotenbeschriftungen aktivieren, indem Sie folgende Eigenschaften konfigurieren: `yarn.node-labels.enabled` und `yarn.node-labels.am.default-node-label-expression`. Weitere Informationen finden Sie unter [Grundlegendes zu Primär-, Kern- und Aufgabenknoten](#).

## Bekannte Probleme

- Niedrigeres Limit für die „Max. Anzahl geöffneter Dateien“ auf älteren AL2-Versionen [in neueren Versionen behoben]. Amazon-EMR-Versionen: `emr-5.30.x`, `emr-5.31.0`, `emr-5.32.0`, `emr-6.0.0`, `emr-6.1.0` und `emr-6.2.0` basieren auf älteren Versionen von Amazon Linux 2 (AL2), die eine niedrigere `ulimit`-Einstellung für „Max. Anzahl geöffneter Dateien“ haben, wenn Amazon-EMR-Cluster mit dem Standard-AMI erstellt werden. Die Amazon-EMR-Versionen `5.30.1`, `5.30.2`, `5.31.1`, `5.32.1`, `6.0.1`, `6.1.1`, `6.2.1`, `5.33.0`, `6.3.0` und höher enthalten einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“. Versionen mit einem niedrigeren Limit für geöffnete Dateien verursachen beim Senden des Spark-Jobs den Fehler „Zu viele offene Dateien“. In den betroffenen Versionen hat das Amazon-EMR-Standard-AMI eine `ulimit`-Standardeinstellung von 4 096 für „Max. Anzahl geöffneter Dateien“, was unter dem Dateilimit von 65 536 im neuesten Amazon-Linux-2-AMI liegt. Die niedrigere `ulimit`-Einstellung für „Max. Anzahl geöffneter Dateien“ führt dazu, dass der Spark-Job fehlschlägt, wenn der Spark-Treiber und der Executor versuchen, mehr als 4 096 Dateien zu öffnen. Um das Problem zu beheben, verfügt Amazon EMR über ein Bootstrap Action (BA)-Skript, das die `ulimit`-Einstellung bei der Cluster-Erstellung anpasst.

Wenn Sie eine ältere Amazon-EMR-Version verwenden, für die dieses Problem nicht dauerhaft behoben ist, können Sie mit der folgenden Problemumgehung das Instance-Controller-`ulimit` explizit auf maximal 65 536 Dateien festlegen.

## Explizit ein ulimit über die Befehlszeile setzen

1. Bearbeiten Sie `/etc/systemd/system/instance-controller.service`, um die folgenden Parameter zum Abschnitt Service hinzuzufügen.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Starten Sie neu InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

## Mithilfe der Bootstrap-Aktion (BA) ein Ulimit festlegen

Sie können auch ein Bootstrap-Aktionsskript (BA) verwenden, um das Ulimit für den Instance-Controller bei der Clustererstellung auf 65 536 Dateien zu konfigurieren.

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

- Die interaktive Spark-Shell PySpark, einschließlich SparkR und Spark-Shell, unterstützt die Verwendung von Docker mit zusätzlichen Bibliotheken nicht.



- Um Python 3 mit Amazon-EMR-Version 6.0.0 zu verwenden, müssen Sie PATH zu `yarn.nodemanager.env-whitelist` hinzufügen.
- Die Funktion Live Long and Process (LLAP) wird nicht unterstützt, wenn Sie den AWS Glue-Datenkatalog als Metastore für Hive verwenden.
- Wenn Sie Amazon EMR 6.0.0 mit Spark- und Docker-Integration verwenden, müssen Sie die Instances in Ihrem Cluster mit demselben Instance-Typ und derselben Anzahl an EBS-Volumes konfigurieren, um Fehler beim Senden eines Spark-Auftrags mit Docker-Laufzeit zu vermeiden.
- [In Amazon EMR 6.0.0 ist der Speichermodus HBase auf Amazon S3 vom HBASE-24286.-Problem betroffen.](#) Der HBase-Master kann nicht initialisiert werden, wenn der Cluster mit vorhandenen S3-Daten erstellt wird.
- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in Amazon-EMR-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie dem Herunterskalieren oder der schrittweisen Übermittlung auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

Workaround:

- SSH als `hadoop`-Benutzer für den führenden Primärknoten des EMR-Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den `hadoop`-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

#### Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen

konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

## 6.0.0 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.2.6	Amazon SageMaker Spark-SDK
<code>emr-ddb</code>	4.14.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	3.0.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.5.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-s3-dist-cp	2.14.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	1.5.0	EMR S3Select-Konnektor
emrfs	2.39.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	3.2.1-amzn-0	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	3.2.1-amzn-0	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	3.2.1-amzn-0	HDFS-Client und -Bibliothek für die Befehlszeile

Komponente	Version	Beschreibung
hadoop-hdfs-namenode	3.2.1-amzn-0	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-hdfs-journalnode	3.2.1-amzn-0	HDFS-Service zum Verwalten des Hadoop-Dateisystem journals auf HA-Clustern.
hadoop-httfs-server	3.2.1-amzn-0	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	3.2.1-amzn-0	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	3.2.1-amzn-0	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	3.2.1-amzn-0	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resource-manager	3.2.1-amzn-0	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	3.2.1-amzn-0	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.

Komponente	Version	Beschreibung
hbase-hmaster	2.2.3	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	2.2.3	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	2.2.3	HBase-Befehlszeilen-Client.
hbase-rest-server	2.2.3	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	2.2.3	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	3.1.2-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	3.1.2-amzn-0	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	3.1.2-amzn-0	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	3.1.2-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	3.1.2-amzn-0	Hive-hbase client.

Komponente	Version	Beschreibung
hive-metastore-server	3.1.2-amzn-0	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	3.1.2-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.5.0-incubating-amzn-1	Inkrementelles Verarbeitungs-Framework zur Stromversorgung der Datenpipeline bei geringer Latenz und hoher Effizienz.
hudi-presto	0.5.0-incubating-amzn-1	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hue-server	4.4.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	1.0.0	Multi-User-Server für Jupyter-Notebooks
livy-server	0.6.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP- und Reverse-Proxy-Server.
mxnet	1.5.1	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.

Komponente	Version	Beschreibung
mariadb-server	5.5.64+	MariaDB-Datenbankserver.
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.1.0	Oozie-Befehlszeilen-Client.
oozie-server	5.1.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	5.0.0-HBase-2.0	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	5.0.0-HBase-2.0	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.230	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.230	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.230	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.

Komponente	Version	Beschreibung
r	3.4.3	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
spark-client	2.4.4	Spark-Befehlszeilen-Clients.
spark-history-server	2.4.4	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.4.4	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.4.4	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
tensorflow	1.14.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.2	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.41+	Apache HTTP-Server.
zeppelin-server	0.9.0-SNAPSHOT	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.



Komponente	Version	Beschreibung
zookeeper-server	3.4.14	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.14	ZooKeeper Befehlszeilen-Cli-ent.

## 6.0.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

### emr-6.0.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-executor	Ändern Sie die Werte in der Datei „ <code>container-executor.cfg</code> “ Datei von Hadoop YARN.
container-log4j	Ändert die Werte in der <code>container-log4j.properties</code> -Datei in Hadoop YARN.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.

Klassifizierungen	Beschreibung
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-env	Ändert die Werte in der HDFS-Umgebung.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.

Klassifizierungen	Beschreibung
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei proto-hive-site.xml von HCatalog.
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Hive.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.

Klassifizierungen	Beschreibung
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.

Klassifizierungen	Beschreibung
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.

Klassifizierungen	Beschreibung
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.
ranger-kms-dbks-site	Ändert die Werte in der dbks-site.xml-Datei von Ranger KMS.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger KMS.
ranger-kms-env	Ändert die Werte in der Ranger KMS-Umgebung.
ranger-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei von Ranger KMS.
ranger-kms-db-ca	Ändert die Werte für die CA-Datei auf S3 für die MySQL SSL-Verbindung mit Ranger KMS.
recordserver-env	Werte in der RecordServer EMR-Umgebung ändern.
recordserver-conf	Ändern Sie die Werte in der Datei RecordServer server.properties von EMR.
recordserver-log4j	Ändern Sie die Werte in der Datei RecordServer log4j.properties von EMR.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.

Klassifizierungen	Beschreibung
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-5.x-Versionen

Dieser Abschnitt enthält Anwendungsversionen, Versionshinweise, Komponentenversionen und Konfigurationsklassifizierungen, die in jeder Amazon-EMR-5.x-Version verfügbar ist.

Wenn Sie einen Cluster starten, können Sie aus verschiedenen Versionen von Amazon EMR wählen. Auf diese Weise können Sie Anwendungsversionen testen und verwenden, die zu den



Kompatibilitätsanforderungen passen. Sie geben die Versionsnummer zusammen mit dem Release-Label an. Versionsbezeichnungen haben die Form `emr-x.x.x`. z. B. `emr-7.1.0`.

Neue Amazon-EMR-Versionen werden über einen Zeitraum von mehreren Tagen in verschiedenen Regionen verfügbar gemacht, beginnend mit der ersten Region am ursprünglichen Veröffentlichungsdatum. Die neueste Version ist in diesem Zeitraum möglicherweise nicht in Ihrer Region verfügbar.

Eine umfassende Tabelle der Anwendungsversionen in jeder Amazon-EMR-5.x-Version finden Sie unter [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#).

## Themen

- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Amazon EMR Version 5.36.2](#)
- [Amazon-EMR-Version 5.36.1](#)
- [Amazon-EMR-Version 5.36.0](#)
- [Amazon-EMR-Version 5.35.0](#)
- [Amazon-EMR-Version 5.34.0](#)
- [Amazon-EMR-Version 5.33.1](#)
- [Amazon-EMR-Version 5.33.0](#)
- [Amazon-EMR-Version 5.32.1](#)
- [Amazon-EMR-Version 5.32.0](#)
- [Amazon-EMR-Version 5.31.1](#)
- [Amazon-EMR-Version 5.31.0](#)
- [Amazon-EMR-Version 5.30.2](#)
- [Amazon-EMR-Version 5.30.1](#)
- [Amazon-EMR-Version 5.30.0](#)
- [Amazon-EMR-Version 5.29.0](#)
- [Amazon-EMR-Version 5.28.1](#)
- [Amazon-EMR-Version 5.28.0](#)
- [Amazon-EMR-Version 5.27.1](#)
- [Amazon-EMR-Version 5.27.0](#)

- [Amazon-EMR-Version 5.26.0](#)
- [Amazon-EMR-Version 5.25.0](#)
- [Amazon-EMR-Version 5.24.1](#)
- [Amazon-EMR-Version 5.24.0](#)
- [Amazon-EMR-Version 5.23.1](#)
- [Amazon-EMR-Version 5.23.0](#)
- [Amazon-EMR-Version 5.22.0](#)
- [Amazon-EMR-Version 5.21.2](#)
- [Amazon-EMR-Version 5.21.1](#)
- [Amazon-EMR-Version 5.21.0](#)
- [Amazon-EMR-Version 5.20.1](#)
- [Amazon-EMR-Version 5.20.0](#)
- [Amazon-EMR-Version 5.19.1](#)
- [Amazon-EMR-Version 5.19.0](#)
- [Amazon-EMR-Version 5.18.1](#)
- [Amazon-EMR-Version 5.18.0](#)
- [Amazon-EMR-Version 5.17.2](#)
- [Amazon-EMR-Version 5.17.1](#)
- [Amazon-EMR-Version 5.17.0](#)
- [Amazon-EMR-Version 5.16.1](#)
- [Amazon-EMR-Version 5.16.0](#)
- [Amazon-EMR-Version 5.15.1](#)
- [Amazon-EMR-Version 5.15.0](#)
- [Amazon-EMR-Version 5.14.2](#)
- [Amazon-EMR-Version 5.14.1](#)
- [Amazon-EMR-Version 5.14.0](#)
- [Amazon-EMR-Version 5.13.1](#)
- [Amazon-EMR-Version 5.13.0](#)
- [Amazon-EMR-Version 5.12.3](#)

- [Amazon-EMR-Version 5.12.2](#)
- [Amazon-EMR-Version 5.12.1](#)
- [Amazon-EMR-Version 5.12.0](#)
- [Amazon-EMR-Version 5.11.4](#)
- [Amazon-EMR-Version 5.11.3](#)
- [Amazon-EMR-Version 5.11.2](#)
- [Amazon-EMR-Version 5.11.1](#)
- [Amazon-EMR-Version 5.11.0](#)
- [Amazon-EMR-Version 5.10.1](#)
- [Amazon-EMR-Version 5.10.0](#)
- [Amazon-EMR-Version 5.9.1](#)
- [Amazon-EMR-Version 5.9.0](#)
- [Amazon-EMR-Version 5.8.3](#)
- [Amazon-EMR-Version 5.8.2](#)
- [Amazon-EMR-Version 5.8.1](#)
- [Amazon-EMR-Version 5.8.0](#)
- [Amazon-EMR-Version 5.7.1](#)
- [Amazon-EMR-Version 5.7.0](#)
- [Amazon-EMR-Version 5.6.1](#)
- [Amazon-EMR-Version 5.6.0](#)
- [Amazon-EMR-Version 5.5.4](#)
- [Amazon-EMR-Version 5.5.3](#)
- [Amazon-EMR-Version 5.5.2](#)
- [Amazon-EMR-Version 5.5.1](#)
- [Amazon-EMR-Version 5.5.0](#)
- [Amazon-EMR-Version 5.4.1](#)
- [Amazon-EMR-Version 5.4.0](#)
- [Amazon-EMR-Version 5.3.2](#)
- [Amazon-EMR-Version 5.3.1](#)

- [Amazon-EMR-Version 5.3.0](#)
- [Amazon-EMR-Version 5.2.3](#)
- [Amazon-EMR-Version 5.2.2](#)
- [Amazon-EMR-Version 5.2.1](#)
- [Amazon-EMR-Version 5.2.0](#)
- [Amazon-EMR-Version 5.1.1](#)
- [Amazon-EMR-Version 5.1.0](#)
- [Amazon-EMR-Version 5.0.3](#)
- [Amazon-EMR-Version 5.0.2](#)
- [Amazon-EMR-Version 5.0.1](#)
- [Amazon-EMR-Version 5.0.0](#)

## Anwendungsversionen in Amazon-EMR-5.x-Versionen (PNG)

Eine umfassende Tabelle mit den in jeder Amazon-EMR-5.x-Version verfügbaren Anwendungsversionen finden Sie unter [Anwendungsversionen in Amazon-EMR-5.x-Versionen](#) in Ihrem Browser.

### Amazon EMR Version 5.36.2

#### 5.36.2 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Delta](#), [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)

- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

### Informationen zur Anwendungsversion

	emr-5.36.2	emr-5.36.1	emr-5.36.0	emr-5.35.0
AWS SDK for Java	1.12.206	1.12.206	1.12.206	1.12.159
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.14.2	1.14.2	1.14.2	1.14.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.13	1.4.13	1.4.13	1.4.13
HCatalog	2.3.9	2.3.9	2.3.9	2.3.9
Hadoop	2.10.1	2.10.1	2.10.1	2.10.1
Hive	2.3.9	2.3.9	2.3.9	2.3.9
Hudi	0.10.1-amzn-1	0.10.1-amzn-1	0.10.1-amzn-1	0.9.0-amzn-2
Hue	4.10.0	4.10.0	4.10.0	4.10.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	2.6.0	2.6.0	2.1.0	2.1.0
JupyterHub	1.4.1	1.4.1	1.4.1	1.4.1

	emr-5.36.2	emr-5.36.1	emr-5.36.0	emr-5.35.0
Livy	0.7.1	0.7.1	0.7.1	0.7.1
MXNet	1.8.0	1.8.0	1.8.0	1.8.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	4.14.3	4.14.3	4.14.3	4.14.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0,267	0,267	0,267	0,266
Spark	2.4.8	2.4.8	2.4.8	2.4.8
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.4.1	2.4.1	2.4.1	2.4.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.10.0	0.10.0	0.10.0	0.10.0
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

## Versionshinweise zu 5.36.2

Die folgenden Versionshinweise enthalten Informationen für Amazon EMR Version 5.36.2. Die Änderungen beziehen sich auf 5.36.1. Informationen zum Zeitplan der Veröffentlichung finden Sie im [Änderungsprotokoll](#).

### Änderungen, Verbesserungen und behobene Probleme

- Diese Version verbessert die Cluster-Scale-Down-Logik, sodass Amazon EMR die Kernknoten nicht unter die HDFS-Replikationsfaktor-Einstellung für den Cluster herunterskaliert.

Diese Verbesserung erfüllt die Anforderungen an die Datenredundanz und verringert die Wahrscheinlichkeit, dass ein Skalierungsvorgang zum Stillstand kommt.

- Diese Version fügt dem Cluster-Skalierungs-Workflow für die Ausführung Presto oder Trino einen neuen Wiederholungsmechanismus hinzu. Diese Verbesserung reduziert das Risiko, dass die Clustergrößenänderung aufgrund eines einzigen fehlgeschlagenen Größenänderungsvorgangs auf unbestimmte Zeit ausgeführt wird. Sie verbessert auch die Clusterauslastung, da Ihr Cluster schneller hoch- und herunterskaliert wird.
- Behebt ein Problem, bei dem Cluster-Scale-Down-Operationen möglicherweise zum Stillstand kommen, während Amazon EMR einen Core-Knoten ordnungsgemäß außer Betrieb nimmt und dieser vor der vollständigen Außerbetriebnahme fehlerhaft wird.
- Verbessert die Stabilität eines Knotens in einem Hochverfügbarkeitscluster mit mehreren Primärknoten, wenn Amazon EMR einen einzelnen Knoten neu startet.
- Optimiert die Protokollverwaltung mit Amazon EMR, das auf Amazon EC2 ausgeführt wird. Infolgedessen können Sie eine leichte Senkung der Speicherkosten für Ihre Cluster-Protokolle feststellen.
- Verbessert die Verwaltung von ZooKeeper Transaktionsprotokolldateien, die auf primären Knoten verwaltet werden, um Szenarien zu minimieren, in denen die Protokolldateien über die Grenzen hinaus wachsen und Clusteroperationen unterbrechen.
- Behebt einen seltenen Fehler, der dazu führen kann, dass ein Hochverfügbarkeitscluster mit mehreren Primärknoten ausfällt, weil er nicht mit dem Yarn kommunizieren kann.  
ResourceManager
- Wenn Sie einen Cluster mit der neuesten Patch-Version von Amazon EMR 5.36 oder höher, 6.6 oder höher oder 7.0 oder höher starten, verwendet Amazon EMR die neueste Version von Amazon Linux 2023 oder Amazon Linux 2 für das standardmäßige Amazon EMR-AMI. Weitere Informationen finden Sie unter [Verwenden des standardmäßigen Amazon-Linux-AMI für Amazon EMR](#).

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 503.0	4,14.343	xxxxxx, 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West),



OsReleaseLabel (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

### 5.36.2 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.4.1	Amazon SageMaker Spark-SDK
<code>emr-ddb</code>	4.16.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.16.0	Praktische Bibliotheken für das Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-kinesis	3.5.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-notebook-env	1.5.0	Conda Env für EMR-Notebooks, das Jupyter Enterprise Gateway enthält
emr-s3-dist-cp	2.21.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	1.7.0	EMR S3Select-Konnektor
emrfs	2,51,0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1,14,2	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
flink-jobmanager-config	1.14.2	Verwaltung von Ressourcen auf EMR-Knoten für Apache JobManager Flink.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.

Komponente	Version	Beschreibung
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.10.1-amzn-4	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.10.1-amzn-4	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.10.1-amzn-4	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.10.1-amzn-4	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-hdfs-journalnode	2.10.1-amzn-4	HDFS-Service zum Verwalten des Hadoop-Dateisystem journals auf HA-Clustern.
hadoop-httpfs-server	2.10.1-amzn-4	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.10.1-amzn-4	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.10.1-amzn-4	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.

Komponente	Version	Beschreibung
hadoop-yarn-nodemanager	2.10.1-amzn-4	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.10.1-amzn-4	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.10.1-amzn-4	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.4.13	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.4.13	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.4.13	HBase-Befehlszeilen-Client.
hbase-rest-server	1.4.13	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.4.13	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.9-amzn-2	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.

Komponente	Version	Beschreibung
hcatalog-server	2.3.9-amzn-2	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.3.9-amzn-2	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.3.9-amzn-2	Hive-Befehlszeilen-Client.
hive-hbase	2.3.9-amzn-2	Hive-hbase client.
hive-metastore-server	2.3.9-amzn-2	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.3.9-amzn-2	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.10.1-amzn-1	Inkrementelles Verarbeitungs-Framework zur Versorgung der Datenpipeline mit geringer Latenz und hoher Effizienz.
hudi-spark	0.10.1-amzn-1	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.
hudi-presto	0.10.1-amzn-1	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.

Komponente	Version	Beschreibung
hue-server	4.10.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	1.4.1	Multi-User-Server für Jupyter-Notebooks
livy-server	0.7.1-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
nginx	1.13.1	nginx [engine x] ist ein HTTP- und Reverse-Proxy-Server.
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.8.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.68	MySQL-Datenbankserver.
nvidia-cuda	11,0,194	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.2.1	Oozie-Befehlszeilen-Client.
oozie-server	5.2.1	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	4.14.3-HBase-1.4	Die Phoenix-Bibliotheken für den Server und den Client

Komponente	Version	Beschreibung
phoenix-query-server	4.14.3-HBase-1.4	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.267-amzn-1	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.267-amzn-1	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.267-amzn-1	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	4.0.2	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	1.2.0	Apache Ranger Key Management System
spark-client	2.4.8-amzn-2	Spark-Befehlszeilen-Clients.
spark-history-server	2.4.8-amzn-2	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.

Komponente	Version	Beschreibung
spark-on-yarn	2.4.8-amzn-2	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.4.8-amzn-2	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	2.4.1	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.2	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.10.0	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.14	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.14	ZooKeeper Befehlszeilen-Client.



## 5.36.2 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Aktionen zur Neukonfiguration treten auf, wenn Sie eine Konfiguration für Instance-Gruppen in einem laufenden Cluster angeben. Amazon EMR initiiert nur Rekonfigurationsaktionen für die Klassifizierungen, die Sie ändern. Weitere Informationen finden Sie unter [Eine Instance-Gruppe in einem laufenden Cluster neu konfigurieren](#).

### emr-5.36.2-Klassifizierungen

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.	Restarts the Resource Manager service.
container-executor	Ändern Sie die Werte in der Datei „ <code>container-executor.cfg</code> “ Datei von Hadoop YARN.	Not available.
container-log4j	Ändert die Werte in der <code>container-log4j.properties</code> -Datei in Hadoop YARN.	Not available.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
		Httpfs, and MapReduce-HistoryServer.
docker-conf	Ändern Sie die Docker-bezogenen Einstellungen.	Not available.
emrfs-site	Ändert die EMRFS-Einstellungen.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegistrator, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.	Restarts Flink history server.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.	Restarts Flink history server.
flink-log4j-session	Ändern Sie die Einstellungen von Flink log4j-session.properties für Kubernetes-/Yarn-Sitzungen.	Not available.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.	Restarts Flink history server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.	Not available.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Ändert die Werte in der HBase-Umgebung.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.	Not available.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.	Should not be reconfigured.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der XML-Datei von HCatalog. proto-hive-site	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.	Restarts Hive WebHCat Server.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.	Restarts Hive WebHCat Server.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.	Restarts Hive WebHCat Server.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.	Not available.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.	Not available.
hive-env	Ändert die Werte in der Hive-Umgebung.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.	Restarts HiveServer2 and HiveMetastore.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log	Not available.
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.	Not available.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.	Not available.
hue-ini	Ändert die Werte in der INI-Datei in Hue.	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.	Restarts Hadoop Httpfs service.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.	Not available.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.	Restarts Hadoop-KMS service.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.	Not available.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Ändern der Werte in der Hudi-Umgebung.	Not available.
hudi-defaults	Ändern Sie die Werte in der hudi-defaults.conf-Datei in Hudi.	Not available.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.	Not available.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.	Not available.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.	Not available.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.	Not available.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.	Restarts Livy Server.
livy-env	Ändert die Werte in der Livy-Umgebung.	Restarts Livy Server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.	Restarts Livy Server.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Ändert die Werte in der Oozie-Umgebung.	Restarts Oozie.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.	Restarts Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.	Not available.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.	Not available.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.	Not available.



Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
pig-env	Ändert die Werte in der Pig-Umgebung.	Not available.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.	Restarts Oozie.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.	Not available.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.	Restarts Presto-Server.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.	Restarts Presto-Server.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.	Not available.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.	Restarts Presto-Server.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.	Not available.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.	Not available.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.	Restarts Presto-Server.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.	Not available.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.	Not available.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.	Not available.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.	Not available.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.	Not available.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.	Not available.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.	Not available.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.	Not available.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.	Not available.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.	Not available.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.	Not available.
ranger-kms-dbks-site	Ändert die Werte in der dbks-site.xml-Datei von Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Ändert die Werte in der Ranger KMS-Umgebung.	Restarts Ranger KMS Server.
ranger-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei von Ranger KMS.	Not available.
ranger-kms-db-ca	Ändert die Werte für die CA-Datei auf S3 für die MySQL SSL-Verbindung mit Ranger KMS.	Not available.
recordserver-env	Werte in der RecordServer EMR-Umgebung ändern.	Restarts EMR record server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
recordserver-conf	Ändern Sie die Werte in der Datei RecordServer server.properties von EMR.	Restarts EMR record server.
recordserver-log4j	Ändern Sie die Werte in der Datei RecordServer log4j.properties von EMR.	Restarts EMR record server.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Ändert die Werte in der Spark-Umgebung.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.	Not available.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.	Not available.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml von Sqoop OraOop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.	Not available.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.	Restarts Oozie.
yarn-env	Ändert die Werte in der YARN-Umgebung.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.	Restarts Zeppelin.
zeppelin-site	Ändern Sie die Konfigurationseinstellungen in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.	Restarts Zookeeper server.

## 5.36.2 Änderungsprotokoll

Änderungsprotokoll für die Version 5.36.2 und die Versionshinweise

Datum	Ereignis	Beschreibung
2024-05-28	Die Bereitstellung ist abgeschlossen	<a href="#">Amazon EMR 5.36.2 vollständig in allen unterstützten Regionen bereitgestellt</a>
28.05.2021	Veröffentlichung von Dokumenten	Erste Veröffentlichung der Versionshinweise zu Amazon EMR 5.36.2
20.05.2024-05	Erstversion	Amazon EMR 5.36.2 wurde zuerst in begrenzten kommerziellen Regionen eingesetzt

## Amazon-EMR-Version 5.36.1

### 5.36.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

## Informationen zur Anwendungsversion

	emr-5.36.1	emr-5.36.0	emr-5.35.0	emr-5.34.0
AWS SDK for Java	1.12.206	1.12.206	1.12.159	1.11.970
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.14.2	1.14.2	1.14.2	1.13.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.13	1.4.13	1.4.13	1.4.13
HCatalog	2.3.9	2.3.9	2.3.9	2.3.8
Hadoop	2.10.1	2.10.1	2.10.1	2.10.1
Hive	2.3.9	2.3.9	2.3.9	2.3.8
Hudi	0.10.1-amzn-1	0.10.1-amzn-1	0.9.0-amzn-2	0.9.0-amzn-0
Hue	4.10.0	4.10.0	4.10.0	4.9.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	2.6.0	2.1.0	2.1.0	2.1.0
JupyterHub	1.4.1	1.4.1	1.4.1	1.4.1
Livy	0.7.1	0.7.1	0.7.1	0.7.1
MXNet	1.8.0	1.8.0	1.8.0	1.8.0

	emr-5.36.1	emr-5.36.0	emr-5.35.0	emr-5.34.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	4.14.3	4.14.3	4.14.3	4.14.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0,267	0,267	0,266	0,261
Spark	2.4.8	2.4.8	2.4.8	2.4.8
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.4.1	2.4.1	2.4.1	2.4.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.10.0	0.10.0	0.10.0	0.10.0
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

## 5.36.1 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon EMR Version 5.36.1. Änderungen beziehen sich auf Version 5.36.0. Informationen zum Zeitplan der Veröffentlichung finden Sie im [Änderungsprotokoll](#).

### Änderungen, Verbesserungen und behobene Probleme

- Amazon EMR Version 5.36.1 bietet Unterstützung für die Archivierung von Protokollen in Amazon S3 während der Cluster-Herunterskalierung. In früheren 5.x-Versionen konnten Sie Protokolldateien nur während der Clusterbeendigung in Amazon S3 archivieren. Die Verbesserung stellt sicher, dass auf dem Cluster generierte Protokolldateien auch nach dem Beenden des



Knotens auf Amazon S3 bestehen bleiben. Weitere Informationen finden Sie unter [Konfigurieren der Cluster-Protokollierung und des Debuggings](#).

- Die Version 5.36.1 verbessert den On-Cluster-Log-Management-Daemon zur Überwachung zusätzlicher Protokollordner in Ihrem EMR-Cluster. Durch diese Verbesserung werden Szenarien mit übermäßiger Festplattenauslastung minimiert.
- Die Version 5.36.1 startet den Log-Management-Daemon auf dem Cluster automatisch neu, wenn er beendet wird. Durch diese Verbesserung wird das Risiko verringert, dass Knoten aufgrund übermäßiger Festplattenauslastung als fehlerhaft erscheinen.
- Die Version 5.36.1 behebt ein Problem, bei dem Amazon-EMR-Daemons auf dem Primärknoten veraltete Metadaten für beendete Instances im Cluster beibehielten. Die Beibehaltung veralteter Daten kann dazu führen, dass die CPU- und Speicherauslastung innerhalb des Clusters unbegrenzt zunimmt, was letztendlich zu Clusterausfällen führen kann.
- Bei Clustern, die mit mehreren Primärknoten gestartet werden, behebt die Version 5.36.1 ein Problem, bei dem ein Amazon-EC2-Hardwarefehler auf einem der Primärknoten zum Ausfall eines zweiten Primärknotens führen und Ihren Cluster instabil machen könnte.
- Für Cluster, die mit Verschlüsselung während der Übertragung konfiguriert sind, unterstützt verwaltete Skalierung jetzt die Datenmenge von Spark Shuffle. Spark-Shuffle-Daten sind Daten, die Spark auf mehrere Partitionen verteilt, um bestimmte Operationen auszuführen. Beim Herunterskalieren ignoriert verwaltete Skalierung die Instances mit Shuffle-Daten. Dadurch werden wiederholte Auftragsversuche und Neuberechnungen verhindert, was sich negativ auf den Preis und die Leistung auswirkt. [Weitere Informationen zu Shuffle-Vorgängen finden Sie im Spark-Programmierhandbuch](#).
- Wenn Sie einen Cluster mit der neuesten Patch-Version von Amazon EMR 5.36 oder höher, 6.6 oder höher oder 7.0 oder höher starten, verwendet Amazon EMR die neueste Version von Amazon Linux 2023 oder Amazon Linux 2 für das standardmäßige Amazon EMR-AMI. Weitere Informationen finden Sie unter [Verwenden des standardmäßigen Amazon-Linux-AMI für Amazon EMR](#).

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 223.0	4,14.336	8. März 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024131,0	4,14.336	14. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 124,0	4,14.336	7. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024109,0	4,14.334	24. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West),



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 218,0	4,14.330	2. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202306,0	4,14.330	22. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada AWS GovCloud (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023116,0	4,14.328	11. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada AWS GovCloud (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023101,0	4,14.327	16. November 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada AWS GovCloud (Zentral), AWS GovCloud (US-West),



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023020,1	4,14.326	7. November 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada AWS GovCloud (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023012,1	4,14.326	26. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada AWS GovCloud (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 926,0	4,14.322	19. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada AWS GovCloud (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)
2.0.2023 906,0	4,14.322	04. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 822.0	4,14.322	30. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral)



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.20230808.0	4,14.320	24. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 727.0	4,14.320	14. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.20230719.0	4,14.320	02. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 628.0	4,14.318	12. Juli 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 612,0	4,14.314	23. Juni 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 404.1	4,14.311	18. April 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral)

### 5.36.1 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version

sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-amzn-*EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen myapp-component der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als 2.2-amzn-2 aufgeführt.

Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1.4.1	Amazon SageMaker Spark-SDK
emr-ddb	4.16.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.16.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.5.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-notebook-env	1.5.0	Conda Env für EMR-Notebooks, das Jupyter Enterprise Gateway enthält
emr-s3-dist-cp	2.21.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	1.7.0	EMR S3Select-Konnektor
emrfs	2,51,0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
flink-client	1,14,2	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
flink-jobmanager-config	1.14.2	Verwaltung von Ressourcen auf EMR-Knoten für Apache JobManager Flink.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.10.1-amzn-4	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.10.1-amzn-4	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.10.1-amzn-4	HDFS-Client und -Bibliothek für die Befehlszeile



Komponente	Version	Beschreibung
hadoop-hdfs-namenode	2.10.1-amzn-4	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-hdfs-journalnode	2.10.1-amzn-4	HDFS-Service zum Verwalten des Hadoop-Dateisystem journals auf HA-Clustern.
hadoop-httfs-server	2.10.1-amzn-4	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.10.1-amzn-4	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.10.1-amzn-4	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.10.1-amzn-4	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resource-manager	2.10.1-amzn-4	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.10.1-amzn-4	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.

Komponente	Version	Beschreibung
hbase-hmaster	1.4.13	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.4.13	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.4.13	HBase-Befehlszeilen-Client.
hbase-rest-server	1.4.13	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.4.13	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.9-amzn-2	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.9-amzn-2	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.3.9-amzn-2	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.3.9-amzn-2	Hive-Befehlszeilen-Client.
hive-hbase	2.3.9-amzn-2	Hive-hbase client.

Komponente	Version	Beschreibung
hive-metastore-server	2.3.9-amzn-2	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.3.9-amzn-2	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.10.1-amzn-1	Inkrementelles Verarbeitungs-Framework zur Versorgung der Datenpipeline mit geringer Latenz und hoher Effizienz.
hudi-spark	0.10.1-amzn-1	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.
hudi-presto	0.10.1-amzn-1	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hue-server	4.10.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	1.4.1	Multi-User-Server für Jupyter-Notebooks
livy-server	0.7.1-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
nginx	1.13.1	nginx [engine x] ist ein HTTP- und Reverse-Proxy-Server.

Komponente	Version	Beschreibung
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.8.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.68	MySQL-Datenbankserver.
nvidia-cuda	11,0,194	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.2.1	Oozie-Befehlszeilen-Client.
oozie-server	5.2.1	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	4.14.3-HBase-1.4	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.14.3-HBase-1.4	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.267-amzn-1	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.267-amzn-1	Service für das Ausführen von Teilen einer Abfrage.

Komponente	Version	Beschreibung
presto-client	0.267-amzn-1	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	4.0.2	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	1.2.0	Apache Ranger Key Management System
spark-client	2.4.8-amzn-2	Spark-Befehlszeilen-Clients.
spark-history-server	2.4.8-amzn-2	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.4.8-amzn-2	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.4.8-amzn-2	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.

Komponente	Version	Beschreibung
tensorflow	2.4.1	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.2	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.10.0	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.14	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.14	ZooKeeper Befehlszeilen-Cli-ent.

### 5.36.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Aktionen zur Neukonfiguration treten auf, wenn Sie eine Konfiguration für Instance-Gruppen in einem laufenden Cluster angeben. Amazon EMR initiiert nur Rekonfigurationsaktionen für die Klassifizierungen, die Sie ändern. Weitere Informationen finden Sie unter [Eine Instance-Gruppe in einem laufenden Cluster neu konfigurieren](#).

## emr-5.36.1-Klassifizierungen

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
capacity-scheduler	Ändert die Werte in der capacity-scheduler.xml-Datei in Hadoop.	Restarts the ResourceManager service.
container-executor	Ändern Sie die Werte in der Datei „container-executor.cfg“ Datei von Hadoop YARN.	Not available.
container-log4j	Ändert die Werte in der container-log4j.properties-Datei in Hadoop YARN.	Not available.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Ändern Sie die Docker-bezogenen Einstellungen.	Not available.
emrfs-site	Ändert die EMRFS-Einstellungen.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
		Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegionserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.	Restarts Flink history server.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.	Restarts Flink history server.
flink-log4j-session	Ändern Sie die Einstellungen von Flink log4j-session.properties für Kubernetes-/Yarn-Sitzungen.	Not available.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.	Restarts Flink history server.



Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.	Not available.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Ändert die Werte in der HBase-Umgebung.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.	Not available.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.	Should not be reconfigured.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei proto-hive-site.xml von HCatalog.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.	Restarts Hive WebHCat Server.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.	Restarts Hive WebHCat Server.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.	Restarts Hive WebHCat Server.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.	Not available.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.	Not available.
hive-env	Ändert die Werte in der Hive-Umgebung.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.	Restarts HiveServer2 and HiveMetastore.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log	Not available.
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.	Not available.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.	Not available.
hue-ini	Ändert die Werte in der INI-Datei in Hue.	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.	Restarts Hadoop Httpfs service.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.	Not available.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.	Restarts Hadoop-KMS service.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.	Not available.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Ändern der Werte in der Hudi-Umgebung.	Not available.
hudi-defaults	Ändern Sie die Werte in der hudi-defaults.conf-Datei in Hudi.	Not available.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.	Not available.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.	Not available.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.	Not available.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.	Not available.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.	Restarts Livy Server.
livy-env	Ändert die Werte in der Livy-Umgebung.	Restarts Livy Server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.	Restarts Livy Server.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Ändert die Werte in der Oozie-Umgebung.	Restarts Oozie.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.	Restarts Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.	Not available.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.	Not available.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
pig-env	Ändert die Werte in der Pig-Umgebung.	Not available.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.	Restarts Oozie.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.	Not available.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.	Restarts Presto-Server.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.	Restarts Presto-Server.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.	Not available.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.	Restarts Presto-Server.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.	Not available.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.	Not available.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.	Restarts Presto-Server.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.	Not available.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.	Not available.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.	Not available.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.	Not available.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.	Not available.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.	Not available.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.	Not available.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.	Not available.



Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.	Not available.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.	Not available.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.	Not available.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.	Not available.
ranger-kms-dbks-site	Ändert die Werte in der dbks-site.xml-Datei von Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Ändert die Werte in der Ranger KMS-Umgebung.	Restarts Ranger KMS Server.
ranger-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei von Ranger KMS.	Not available.
ranger-kms-db-ca	Ändert die Werte für die CA-Datei auf S3 für die MySQL SSL-Verbindung mit Ranger KMS.	Not available.
recordserver-env	Werte in der RecordServer EMR-Umgebung ändern.	Restarts EMR record server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
recordserver-conf	Ändern Sie die Werte in der Datei RecordServer server.properties von EMR.	Restarts EMR record server.
recordserver-log4j	Ändern Sie die Werte in der Datei RecordServer log4j.properties von EMR.	Restarts EMR record server.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Ändert die Werte in der Spark-Umgebung.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.	Not available.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.	Not available.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml von Sqoop OraOop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.	Not available.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.	Restarts Oozie.
yarn-env	Ändert die Werte in der YARN-Umgebung.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.	Restarts Zeppelin.
zeppelin-site	Ändern Sie die Konfigurationseinstellungen in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.	Restarts Zookeeper server.

## 5.36.1 Änderungsprotokoll

### Änderungsprotokoll für Version 5.36.1 und Versionshinweise

Datum	Ereignis	Beschreibung
2023-07-26	Aktualisierung	Neue Betriebssystem-Release-Labels 2.0.20230612.0 und 2.0.20230628.0 .
2023-05-25	Die Bereitstellung ist abgeschlossen	<a href="#">Amazon EMR 5.36.1 vollständig in allen unterstützten Regionen bereitgestellt</a>
2023-05-09	Veröffentlichung von Dokumenten	Erste Veröffentlichung der Versionshinweise zu Amazon EMR 5.36.1
2023-05-04	Erstversion	Amazon EMR 5.36.1 wurde zuerst in begrenzten kommerziellen Regionen bereitgestellt

## Amazon-EMR-Version 5.36.0

### 5.36.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

### Informationen zur Anwendungsversion

	emr-5.36.0	emr-5.35.0	emr-5.34.0	emr-5.33.1
AWS SDK for Java	1.12.206	1.12.159	1.11.970	1,11,970
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.14.2	1.14.2	1.13.1	1.12.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.13	1.4.13	1.4.13	1.4.13
HCatalog	2.3.9	2.3.9	2.3.8	2.3.7
Hadoop	2.10.1	2.10.1	2.10.1	2.10.1
Hive	2.3.9	2.3.9	2.3.8	2.3.7
Hudi	0.10.1-amzn-1	0.9.0-amzn-2	0.9.0-amzn-0	0.7.0-amzn-1
Hue	4.10.0	4.10.0	4.9.0	4.9.0
Iceberg	-	-	-	-

	emr-5.36.0	emr-5.35.0	emr-5.34.0	emr-5.33.1
JupyterEnterpriseGateway	2.1.0	2.1.0	2.1.0	2.1.0
JupyterHub	1.4.1	1.4.1	1.4.1	1.2.2
Livy	0.7.1	0.7.1	0.7.1	0.7.0
MXNet	1.8.0	1.8.0	1.8.0	1.7.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.1	5.2.1	5.2.1	5.2.0
Phoenix	4.14.3	4.14.3	4.14.3	4.14.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0,267	0,266	0,261	0,245,1
Spark	2.4.8	2.4.8	2.4.8	2.4.7
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.4.1	2.4.1	2.4.1	2.4.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.10.0	0.10.0	0.10.0	0.9.0
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

## 5.36.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.36.0. Änderungen beziehen sich auf Version 5.35.0.

Erste Version: 15. Juni 2022

## Neue Features

- Amazon-EMR-Version 5.36.0 bietet Unterstützung für Data Definition Language (DDL) mit Apache Spark auf Apache Ranger-fähigen Clustern. Auf diese Weise können Sie Apache Ranger für die Verwaltung des Zugriffs für Operationen wie das Erstellen, Ändern und Löschen von Datenbanken und Tabellen aus einem Amazon-EMR-Cluster verwenden.
- Wenn Sie einen Cluster mit der neuesten Patch-Version von Amazon EMR 5.36 oder höher, 6.6 oder höher oder 7.0 oder höher starten, verwendet Amazon EMR die neueste Version von Amazon Linux 2023 oder Amazon Linux 2 für das standardmäßige Amazon EMR-AMI. Weitere Informationen finden Sie unter [Verwenden des standardmäßigen Amazon-Linux-AMI für Amazon EMR](#).

### Note

Diese Version erhält keine automatischen AMI-Updates mehr, da sie von einer weiteren Patch-Version abgelöst wurde. Die Patch-Version wird durch die Zahl nach dem zweiten Dezimalpunkt (6.8.**1**) gekennzeichnet. Um zu sehen, ob Sie die neueste Patch-Version verwenden, überprüfen Sie die verfügbaren Versionen im [Versionshandbuch](#) oder überprüfen Sie die Amazon-EMR-Versions-Dropdownliste, wenn Sie einen Cluster in der Konsole erstellen, oder verwenden Sie die [ListReleaseLabels](#)-API- oder [list-release-labels](#)-CLI-Aktion. Um aktuelle Informationen über Neuerscheinungen erhalten, abonnieren Sie den RSS-Feed auf der Seite [Was ist neu?](#).

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 504.1	4,14.313	16. Mai 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 418,0	4,14.311	3. Mai 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 404.1	4,14.311	18. April 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)
2.0.2023 404.0	4,14.311	10. April 2023	USA Ost (Nord-Virginia), Europa (Paris)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 320.0	4,14.309	30. März 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202307,0	4,14.305	15. März 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202307,0	4,14.304	22. Februar 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023119.1	4,14.301	3. Februar 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022 210.1	4.14.301	22. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022103.3	4,14.296	5. Dezember 2022	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022004,0	4,14.294	02. November 2022	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022 912.1	4,14.291	7. Oktober 2022	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)
2.0.2022 719.0	4,14.287	10. August 2022	USA West (Nordkalifornien), Europa (Paris), Europa (Stockholm), Europa (Frankfurt), Asien-Pazifik (Mumbai), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022 426,0	4,14.281	14. Juni 2022	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

### Änderungen, Verbesserungen und behobene Probleme

- Amazon-EMR-5.36.0-Upgrades unterstützen jetzt: aws-sdk 1.12.206, Hadoop 2.10.1-amzn-4, Hive 2.3.9-amzn-2, Hudi 0.10.1-amzn-1, Spark 2.4.8-amzn-2, Presto 0.267-amzn-1, Amazon-Glue-Konnektor 1.18.0, EMRFS 2.51.0.

## Bekannte Probleme

- Wenn Sie Spark mit der Formatierung des Hive-Partitionsstandorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den Amazon-EMR-Versionen 5.30.0 bis 5.36.0 und 6.2.0 bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:
  - Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.
  - Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist `s3://bucket/table/p=a` ein Präfix von `s3://bucket/table/p=a b`.
  - Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen UTF-8-Wert, der kleiner als das `/`-Zeichen (U+002F) ist. Beispielsweise fällt das Leerzeichen (U+0020), das in `s3://bucket/table/p=a b` zwischen `a` und `b` vorkommt, in diese Kategorie. Beachten Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: `!"#$%&'()*+,-.` Weitere Informationen finden Sie unter [UTF-8-Kodierungstabelle und Unicode-Zeichen](#).

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.

- Bei den Amazon-EMR-Versionen 5.36.0 und 6.6.0 bis 6.9.0 kann es bei `SecretAgent`- und `RecordServer`-Servicekomponenten aufgrund einer falschen Konfiguration des Dateinamenmusters in den `Log4j2`-Eigenschaften zu einem Verlust von Protokolldaten kommen. Die falsche Konfiguration führt dazu, dass die Komponenten nur eine Protokolldatei pro Tag generieren. Wenn die Rotationsstrategie angewendet wird, überschreibt sie die vorhandene Datei, anstatt wie erwartet eine neue Protokolldatei zu generieren. Um das Problem zu umgehen, verwenden Sie eine Bootstrap-Aktion, um jede Stunde Protokolldateien zu generieren, und fügen Sie eine automatisch inkrementierte Ganzzahl an den Dateinamen an, um die Rotation zu handhaben.

Verwenden Sie für die Versionen Amazon EMR 6.6.0 bis 6.9.0 die folgende Bootstrap-Aktion, wenn Sie einen Cluster starten.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-6x/replace-puppet.sh,Args=[]"
```

Verwenden Sie für Amazon EMR 5.36.0 die folgende Bootstrap-Aktion, wenn Sie einen Cluster starten.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-5x/replace-puppet.sh,Args=[]"
```

### 5.36.0 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.4.1	Amazon SageMaker Spark-SDK
<code>emr-ddb</code>	4.16.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.16.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.5.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-notebook-env	1.5.0	Conda Env für EMR-Notebooks, das Jupyter Enterprise Gateway enthält
emr-s3-dist-cp	2.21.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	1.7.0	EMR S3Select-Konnektor
emrfs	2,51,0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1,14,2	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
flink-jobmanager-config	1.14.2	Verwaltung von Ressourcen auf EMR-Knoten für Apache JobManager Flink.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.

Komponente	Version	Beschreibung
hadoop-client	2.10.1-amzn-4	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.10.1-amzn-4	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.10.1-amzn-4	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.10.1-amzn-4	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-hdfs-journalnode	2.10.1-amzn-4	HDFS-Service zum Verwalten des Hadoop-Dateisystem journals auf HA-Clustern.
hadoop-httpfs-server	2.10.1-amzn-4	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.10.1-amzn-4	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.10.1-amzn-4	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.10.1-amzn-4	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.

Komponente	Version	Beschreibung
hadoop-yarn-resourcemanager	2.10.1-amzn-4	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.10.1-amzn-4	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.4.13	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.4.13	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.4.13	HBase-Befehlszeilen-Client.
hbase-rest-server	1.4.13	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.4.13	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.9-amzn-2	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.9-amzn-2	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).



Komponente	Version	Beschreibung
hcatalog-webhcat-server	2.3.9-amzn-2	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.3.9-amzn-2	Hive-Befehlszeilen-Client.
hive-hbase	2.3.9-amzn-2	Hive-hbase client.
hive-metastore-server	2.3.9-amzn-2	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.3.9-amzn-2	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.10.1-amzn-1	Inkrementelles Verarbeitungs-Framework zur Versorgung der Datenpipeline mit geringer Latenz und hoher Effizienz.
hudi-spark	0.10.1-amzn-1	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.
hudi-presto	0.10.1-amzn-1	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hue-server	4.10.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	1.4.1	Multi-User-Server für Jupyter-Notebooks

Komponente	Version	Beschreibung
livy-server	0.7.1-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
nginx	1.13.1	nginx [engine x] ist ein HTTP- und Reverse-Proxy-Server.
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.8.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.68	MySQL-Datenbankserver.
nvidia-cuda	11,0,194	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.2.1	Oozie-Befehlszeilen-Client.
oozie-server	5.2.1	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	4.14.3-HBase-1.4	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.14.3-HBase-1.4	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.

Komponente	Version	Beschreibung
presto-coordinator	0.267-amzn-1	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.267-amzn-1	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.267-amzn-1	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	4.0.2	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	1.2.0	Apache Ranger Key Management System
spark-client	2.4.8-amzn-2	Spark-Befehlszeilen-Clients.
spark-history-server	2.4.8-amzn-2	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.4.8-amzn-2	In-Memory-Ausführungs-Engine für YARN.

Komponente	Version	Beschreibung
spark-yarn-slave	2.4.8-amzn-2	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	2.4.1	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.2	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.10.0	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.14	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.14	ZooKeeper Befehlszeilen-Client.

## 5.36.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Aktionen zur Neukonfiguration treten auf, wenn Sie eine Konfiguration für Instance-Gruppen in einem laufenden Cluster angeben. Amazon EMR initiiert nur Rekonfigurationsaktionen für die Klassifizierungen, die Sie ändern. Weitere Informationen finden Sie unter [Eine Instance-Gruppe in einem laufenden Cluster neu konfigurieren](#).

### emr-5.36.0-Klassifizierungen

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.	Restarts the Resource Manager service.
container-executor	Ändern Sie die Werte in der Datei „ <code>container-executor.cfg</code> “ Datei von Hadoop YARN.	Not available.
container-log4j	Ändert die Werte in der <code>container-log4j.properties</code> -Datei in Hadoop YARN.	Not available.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
		Httpfs, and MapReduce-HistoryServer.
docker-conf	Ändern Sie die Docker-bezogenen Einstellungen.	Not available.
emrfs-site	Ändert die EMRFS-Einstellungen.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegistrator, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.	Restarts Flink history server.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.	Restarts Flink history server.
flink-log4j-session	Ändern Sie die Einstellungen von Flink log4j-session.properties für Kubernetes-/Yarn-Sitzungen.	Not available.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.	Restarts Flink history server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.	Not available.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Ändert die Werte in der HBase-Umgebung.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.	Not available.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.	Should not be reconfigured.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.



Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei proto-hive-site.xml von HCatalog.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.	Restarts Hive WebHCat Server.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.	Restarts Hive WebHCat Server.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.	Restarts Hive WebHCat Server.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.	Not available.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.	Not available.
hive-env	Ändert die Werte in der Hive-Umgebung.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.	Restarts HiveServer2 and HiveMetastore.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log	Not available.
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.	Not available.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.	Not available.
hue-ini	Ändert die Werte in der INI-Datei in Hue.	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.	Restarts Hadoop Httpfs service.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.	Not available.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.	Restarts Hadoop-KMS service.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.	Not available.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Ändern der Werte in der Hudi-Umgebung.	Not available.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.	Not available.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.	Not available.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.	Not available.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.	Not available.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.	Restarts Livy Server.
livy-env	Ändert die Werte in der Livy-Umgebung.	Restarts Livy Server.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.	Restarts Livy Server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Ändert die Werte in der Oozie-Umgebung.	Restarts Oozie.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.	Restarts Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.	Not available.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.	Not available.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.	Not available.
pig-env	Ändert die Werte in der Pig-Umgebung.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.	Restarts Oozie.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.	Not available.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.	Restarts Presto-Server.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.	Restarts Presto-Server.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.	Not available.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.	Restarts Presto-Server.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.	Not available.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.	Not available.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.	Not available.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.	Restarts Presto-Server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.	Not available.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.	Not available.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.	Not available.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.	Not available.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.	Not available.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.	Not available.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.	Not available.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.	Not available.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.	Not available.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.	Not available.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.	Not available.
ranger-kms-dbks-site	Ändert die Werte in der dbks-site.xml-Datei von Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Ändert die Werte in der Ranger KMS-Umgebung.	Restarts Ranger KMS Server.
ranger-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei von Ranger KMS.	Not available.
ranger-kms-db-ca	Ändert die Werte für die CA-Datei auf S3 für die MySQL SSL-Verbindung mit Ranger KMS.	Not available.
recordserver-env	Werte in der RecordServer EMR-Umgebung ändern.	Restarts EMR record server.
recordserver-conf	Ändern Sie die Werte in der Datei RecordServer server.properties von EMR.	Restarts EMR record server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
recordserver-log4j	Ändern Sie die Werte in der Datei RecordServer log4j.properties von EMR.	Restarts EMR record server.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Ändert die Werte in der Spark-Umgebung.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.	Not available.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.	Not available.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml von Sqoop OraOop.	Not available.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.	Not available.



Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.	Restarts Oozie.
yarn-env	Ändert die Werte in der YARN-Umgebung.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.	Restarts Zeppelin.
zeppelin-site	Ändern Sie die Konfigurationseinstellungen in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.	Restarts Zookeeper server.

## Amazon-EMR-Version 5.35.0

### 5.35.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

#### Informationen zur Anwendungsversion

	emr-5.35.0	emr-5.34.0	emr-5.33.1	emr-5.33.0
AWS SDK for Java	1.12.159	1.11.970	1,11,970	1,11,970
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.14.2	1.13.1	1.12.1	1.12.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2

	emr-5.35.0	emr-5.34.0	emr-5.33.1	emr-5.33.0
HBase	1.4.13	1.4.13	1.4.13	1.4.13
HCatalog	2.3.9	2.3.8	2.3.7	2.3.7
Hadoop	2.10.1	2.10.1	2.10.1	2.10.1
Hive	2.3.9	2.3.8	2.3.7	2.3.7
Hudi	0.9.0-amzn-2	0.9.0-amzn-0	0.7.0-amzn-1	0.7.0-amzn-1
Hue	4.10.0	4.9.0	4.9.0	4.9.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	2.1.0	2.1.0	2.1.0	2.1.0
JupyterHub	1.4.1	1.4.1	1.2.2	1.2.2
Livy	0.7.1	0.7.1	0.7.0	0.7.0
MXNet	1.8.0	1.8.0	1.7.0	1.7.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.1	5.2.1	5.2.0	5.2.0
Phoenix	4.14.3	4.14.3	4.14.3	4.14.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0,266	0,261	0,245,1	0,245,1
Spark	2.4.8	2.4.8	2.4.7	2.4.7
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.4.1	2.4.1	2.4.1	2.4.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2

	emr-5.35.0	emr-5.34.0	emr-5.33.1	emr-5.33.0
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.10.0	0.10.0	0.9.0	0.9.0
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

## 5.35.0 Versionshinweise

Dies ist der Versionshinweis zu Amazon-EMR-Version 5.35.0.

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.35.0. Änderungen beziehen sich auf Version 5.34.0.

Erste Version: 30. März 2022

### Neue Features

- Amazon-EMR-Anwendungen der Version 5.35, die Log4j 1.x und Log4j 2.x verwenden, werden auf Log4j 1.2.17 (oder höher) bzw. Log4j 2.17.1 (oder höher) aktualisiert und erfordern keine Bootstrap-Aktionen zur Minderung der CVE-Probleme in früheren Versionen. Siehe [Ansatz zur Minderung von CVE-2021-44228](#).

### Änderungen, Verbesserungen und behobene Probleme

#### Änderungen bei Flink

Änderungstyp	Beschreibung
Upgrades	<ul style="list-style-type: none"> <li>• Aktualisieren Sie die Flink-Version auf 1.14.2.</li> <li>• log4j wurde auf 2.17.1 aktualisiert.</li> </ul>

## Hadoop-Änderungen

Änderungstyp	Beschreibung
Hadoop-Open-Source-Backports seit EMR 5.34.0	<ul style="list-style-type: none"> <li>• <a href="#">YARN-10438</a>: Behandle eine Container-ID mit Null in <code>clientRMService# () getContainerReport</code></li> <li>• <a href="#">YARN-7266</a>: <a href="#">Threads für den Timeline-Server-Event-Handler sind gesperrt</a></li> <li>• <a href="#">YARN-10438</a>: ATS 1.5 kann nicht gestartet werden, wenn Dateien beschädigt sind oder fehlen <code>RollingLevelDb</code></li> <li>• <a href="#">HADOOP-13500</a>: Synchronisieren der Iteration des Objekts mit den Konfigurationseigenschaften</li> <li>• <a href="#">YARN-10651</a>: Ist bei eingeschaltetem NPE abgestürzt. <code>CapacityScheduler AbstractYarnScheduler updateNodeResource()</code></li> <li>• <a href="#">HDFS-12221</a>: Ersetzen Sie Xerces in <code>XmlEditsVisitor</code></li> <li>• <a href="#">HDFS-16410</a>: Unsicheres Einlesen von XML <code>OfflineEditsXMLLoader</code></li> </ul>
Hadoop-Änderungen und Korrekturen	<ul style="list-style-type: none"> <li>• Tomcat, das in KMS und HttpFS verwendet wird, wurde auf 8.5.75 aktualisiert</li> <li>• In <code>FileSystemOptimizedCommitter Version 2</code> wurde die Erfolgsmarkierung in den <code>CommitJob</code>-Ausgabepfad geschrieben, der bei der Erstellung des Committers definiert wurde. Da sich die Ausgabepfade von <code>CommitJob</code> und Aufgabenebene unterscheiden können, wurde der Pfad korrigiert, sodass er den in den Manifestdateien definierten Pfad verwendet. Bei Hive-Auftagen führt dies dazu, dass die Erfolgsmarkierung korrekt geschrieben wird, wenn</li> </ul>

Änderungstyp	Beschreibung
	Operationen wie dynamische Partition oder UNION ALL ausgeführt werden.

## Änderungen bei Hive

Änderungstyp	Beschreibung
Hive wurde auf die <a href="#">Open-Source-Version 2.3.9</a> aktualisiert, einschließlich dieser JIRA-Korrekturen	<ul style="list-style-type: none"> <li>• <a href="#">HIVE-17155</a>: findConfFile () in HiveConf .java hat einige Probleme mit dem Conf-Pfad</li> <li>• <a href="#">HIVE-24797</a>: Deaktiviert die Validierung von Standardwerten beim Parsen von Avro-Schemas</li> <li>• <a href="#">HIVE-21563</a>: Verbessern Sie die Leistung von Table#, indem Sie Once deaktivieren getEmptyTable registerAllFunctions</li> <li>• <a href="#">HIVE-18147</a>: Tests können mit java.net fehlschlagen. BindException: Adresse wird bereits verwendet</li> <li>• <a href="#">HIVE-24608</a>: Wechseln Sie im HMS-Client für Hive 2.3.x zurück zu get_table</li> <li>• <a href="#">HIVE-21200</a>: Vektorisierung — Die Datumsspalte wirft java.lang aus. UnsupportedOperationException für Parkett</li> <li>• <a href="#">HIVE-19228</a>: Entfernen Sie die Verwendung von commons-httpclient 3.x</li> </ul>
Hive-Open-Source-Backports seit EMR 5.34.0	<ul style="list-style-type: none"> <li>• <a href="#">HIVE-19990</a>: Eine Abfrage mit einem Intervall literal in der Join-Bedingung schlägt fehl</li> <li>• <a href="#">HIVE-25824</a>: Aktualisieren Sie Branch-2.3 auf log4j 2.17.0</li> <li>• <a href="#">TEZ-4062</a>: Die Planung spekulativer Versuche sollte abgebrochen werden, wenn die Aufgabe abgeschlossen ist</li> </ul>

Änderungstyp	Beschreibung
	<ul style="list-style-type: none"> <li>• <a href="#">TEZ-4108</a>: Rennbedingungen NullPointerException während der spekulativen Ausführung</li> <li>• <a href="#">TEZ-3918</a>: Das Setzen von <code>tez.task.log.level</code> funktioniert nicht</li> </ul>
Hive-Upgrades und -Korrekturen	<ul style="list-style-type: none"> <li>• Aktualisieren Sie die Log4j-Version auf 2.17.1</li> <li>• Auf ORC-Version 1.4.3 upgraden</li> <li>• Der Deadlock aufgrund eines fehlerhaften Thread-Ins wurde behoben ShuffleScheduler</li> </ul>
Neue Features	<ul style="list-style-type: none"> <li>• Feature zum Drucken von Hive-Abfragen in AM-Protokollen hinzugefügt. Diese ist standardmäßig deaktiviert. Flag/Konf.: <code>tez.am.emr.print.hive.query.in.log</code> Status (Standard): FALSE.</li> </ul>

## Oozie-Änderungen

Änderungstyp	Beschreibung
Oozie-Open-Source-Backports seit EMR 5.34.0	<ul style="list-style-type: none"> <li>• <a href="#">OOZIE-3652: Der Oozie-Launcher</a> sollte die Verzeichnisliste erneut versuchen, wenn dies der Fall ist NoSuchFileException</li> </ul>

## Pig-Änderungen

Änderungstyp	Beschreibung
Upgrades	<ul style="list-style-type: none"> <li>• log4j wurde auf 1.2.17 aktualisiert.</li> </ul>

## Bekannte Probleme

- Wenn Sie Spark mit der Formatierung des Hive-Partitionsstandorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den Amazon-EMR-Versionen 5.30.0 bis 5.36.0 und 6.2.0 bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:
  - Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.
  - Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist `s3://bucket/table/p=a` ein Präfix von `s3://bucket/table/p=a b`.
  - Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen UTF-8-Wert, der kleiner als das /-Zeichen (U+002F) ist. Beispielsweise fällt das Leerzeichen (U+0020), das in `s3://bucket/table/p=a b` zwischen a und b vorkommt, in diese Kategorie. Beachten Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: `!"#$%&'()*+,-.` Weitere Informationen finden Sie unter [UTF-8-Kodierungstabelle und Unicode-Zeichen](#).

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.

## 5.35.0 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.



Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1.4.1	Amazon SageMaker Spark-SDK
emr-ddb	4.16.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.15.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.5.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-notebook-env	1.5.0	Conda Env für EMR-Notebooks, das Jupyter Enterprise Gateway enthält
emr-s3-dist-cp	2.20.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	1.7.0	EMR S3Select-Konnektor
emrfs	2,49,0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1,14,2	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
flink-jobmanager-config	1.14.2	Verwaltung von Ressourcen auf EMR-Knoten für Apache JobManager Flink.

Komponente	Version	Beschreibung
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.10.1-amzn-3	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.10.1-amzn-3	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.10.1-amzn-3	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.10.1-amzn-3	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-hdfs-journalnode	2.10.1-amzn-3	HDFS-Service zum Verwalten des Hadoop-Dateisystem journals auf HA-Clustern.

Komponente	Version	Beschreibung
hadoop-httfs-server	2.10.1-amzn-3	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.10.1-amzn-3	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.10.1-amzn-3	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.10.1-amzn-3	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.10.1-amzn-3	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.10.1-amzn-3	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.4.13	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.4.13	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.

Komponente	Version	Beschreibung
hbase-client	1.4.13	HBase-Befehlszeilen-Client.
hbase-rest-server	1.4.13	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.4.13	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.9-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.9-amzn-0	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.3.9-amzn-0	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.3.9-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.3.9-amzn-0	Hive-hbase client.
hive-metastore-server	2.3.9-amzn-0	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.3.9-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.

Komponente	Version	Beschreibung
hudi	0.9.0-amzn-2	Inkrementelles Verarbeitungs-Framework zur Versorgung der Datenpipeline mit geringer Latenz und hoher Effizienz.
hudi-spark	0.9.0-amzn-2	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.
hudi-presto	0.9.0-amzn-2	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hue-server	4.10.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	1.4.1	Multi-User-Server für Jupyter-Notebooks
livy-server	0.7.1-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
nginx	1.13.1	nginx [engine x] ist ein HTTP- und Reverse-Proxy-Server.
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.8.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.68	MySQL-Datenbankserver.
nvidia-cuda	10,1243	Nvidia-Treiber und Cuda-Tool kit

Komponente	Version	Beschreibung
oozie-client	5.2.1	Oozie-Befehlszeilen-Client.
oozie-server	5.2.1	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	4.14.3-HBase-1.4	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.14.3-HBase-1.4	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.266-amzn-0	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.266-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.266-amzn-0	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.

Komponente	Version	Beschreibung
r	4.0.2	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	1.2.0	Apache Ranger Key Management System
spark-client	2.4.8-amzn-1	Spark-Befehlszeilen-Clients.
spark-history-server	2.4.8-amzn-1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.4.8-amzn-1	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.4.8-amzn-1	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	2.4.1	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.2	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.

Komponente	Version	Beschreibung
zeppelin-server	0.10.0	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.14	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.14	ZooKeeper Befehlszeilen-Cli-ent.

### 5.35.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Aktionen zur Neukonfiguration treten auf, wenn Sie eine Konfiguration für Instance-Gruppen in einem laufenden Cluster angeben. Amazon EMR initiiert nur Rekonfigurationsaktionen für die Klassifizierungen, die Sie ändern. Weitere Informationen finden Sie unter [Eine Instance-Gruppe in einem laufenden Cluster neu konfigurieren](#).

#### emr-5.35.0-Klassifizierungen

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.	Restarts the Resource Manager service.



Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
container-executor	Ändern Sie die Werte in der Datei „container-executor.cfg“ Datei von Hadoop YARN.	Not available.
container-log4j	Ändert die Werte in der container-log4j.properties-Datei in Hadoop YARN.	Not available.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Ändern Sie die Docker-bezogenen Einstellungen.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
emrfs-site	Ändert die EMRFS-Einstellungen.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegistrator, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.	Restarts Flink history server.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.	Restarts Flink history server.
flink-log4j-session	Ändern Sie die Einstellungen von Flink log4j-session.properties für Kubernetes-/Yarn-Sitzungen.	Not available.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.	Restarts Flink history server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.	Not available.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Ändert die Werte in der HBase-Umgebung.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.	Not available.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.	Should not be reconfigured.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei proto-hive-site.xml von HCatalog.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.	Restarts Hive WebHCat Server.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.	Restarts Hive WebHCat Server.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.	Restarts Hive WebHCat Server.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.	Not available.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.	Not available.
hive-env	Ändert die Werte in der Hive-Umgebung.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.	Restarts HiveServer2 and HiveMetastore.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log	Not available.
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.	Not available.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.	Not available.
hue-ini	Ändert die Werte in der INI-Datei in Hue.	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.	Restarts Hadoop Httpfs service.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.	Not available.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.	Restarts Hadoop-KMS service.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.	Not available.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Ändern der Werte in der Hudi-Umgebung.	Not available.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.	Not available.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.	Not available.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.	Not available.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.	Not available.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.	Restarts Livy Server.
livy-env	Ändert die Werte in der Livy-Umgebung.	Restarts Livy Server.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.	Restarts Livy Server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
mapred-env	Ändern Sie die Werte in der MapReduce Anwendung sumgebung.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Ändert die Werte in der Oozie-Umgebung.	Restarts Oozie.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.	Restarts Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.	Not available.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.	Not available.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.	Not available.
pig-env	Ändert die Werte in der Pig-Umgebung.	Not available.



Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.	Restarts Oozie.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.	Not available.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.	Restarts Presto-Server.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.	Restarts Presto-Server.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.	Not available.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.	Restarts Presto-Server.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.	Not available.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.	Not available.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.	Not available.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.	Restarts Presto-Server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.	Not available.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.	Not available.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.	Not available.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.	Not available.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.	Not available.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.	Not available.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.	Not available.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.	Not available.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.	Not available.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.	Not available.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.	Not available.
ranger-kms-dbks-site	Ändert die Werte in der dbks-site.xml-Datei von Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Ändert die Werte in der Ranger KMS-Umgebung.	Restarts Ranger KMS Server.
ranger-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei von Ranger KMS.	Not available.
ranger-kms-db-ca	Ändert die Werte für die CA-Datei auf S3 für die MySQL SSL-Verbindung mit Ranger KMS.	Not available.
recordserver-env	Werte in der RecordServer EMR-Umgebung ändern.	Restarts EMR record server.
recordserver-conf	Ändern Sie die Werte in der Datei RecordServer server.properties von EMR.	Restarts EMR record server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
recordserver-log4j	Ändern Sie die Werte in der Datei RecordServer log4j.properties von EMR.	Restarts EMR record server.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Ändert die Werte in der Spark-Umgebung.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.	Not available.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.	Not available.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml von Sqoop OraOop.	Not available.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.	Restarts Oozie.
yarn-env	Ändert die Werte in der YARN-Umgebung.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.	Restarts Zeppelin.
zeppelin-site	Ändern Sie die Konfigurationseinstellungen in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.	Restarts Zookeeper server.

## Amazon-EMR-Version 5.34.0

### 5.34.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

#### Informationen zur Anwendungsversion

	emr-5.34.0	emr-5.33.1	emr-5.33.0	emr-5.32.1
AWS SDK for Java	1.11.970	1,11,970	1,11,970	1,11,890
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.13.1	1.12.1	1.12.1	1.11.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2

	emr-5.34.0	emr-5.33.1	emr-5.33.0	emr-5.32.1
HBase	1.4.13	1.4.13	1.4.13	1.4.13
HCatalog	2.3.8	2.3.7	2.3.7	2.3.7
Hadoop	2.10.1	2.10.1	2.10.1	2.10.1
Hive	2.3.8	2.3.7	2.3.7	2.3.7
Hudi	0.9.0-amzn-0	0.7.0-amzn-1	0.7.0-amzn-1	0.6.0-amzn-0
Hue	4.9.0	4.9.0	4.9.0	4.8.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	2.1.0	2.1.0	2.1.0	2.1.0
JupyterHub	1.4.1	1.2.2	1.2.2	1.1.0
Livy	0.7.1	0.7.0	0.7.0	0.7.0
MXNet	1.8.0	1.7.0	1.7.0	1.7.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.1	5.2.0	5.2.0	5.2.0
Phoenix	4.14.3	4.14.3	4.14.3	4.14.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0,261	0,245,1	0,245,1	0,240,1
Spark	2.4.8	2.4.7	2.4.7	2.4.7
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.4.1	2.4.1	2.4.1	2.3.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2

	emr-5.34.0	emr-5.33.1	emr-5.33.0	emr-5.32.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.10.0	0.9.0	0.9.0	0.8.2
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

## 5.34.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.34.0. Änderungen beziehen sich auf Version 5.33.1.

Erste Version: 20. Januar 2022

Aktualisiertes Veröffentlichungsdatum: 21. März 2022

### Neue Features

- [Verwaltete Skalierung] Spark Shuffle Data Managed Scaling Optimization – Für Amazon-EMR-Versionen 5.34.0 und höher sowie EMR-Versionen 6.4.0 und höher unterstützt verwaltete Skalierung jetzt Spark Shuffle Data (Daten, die Spark partitionsübergreifend verteilt, um bestimmte Operationen auszuführen). Weitere Informationen zu Shuffle-Vorgängen finden Sie unter [Nutzung von Managed Scaling in Amazon EMR](#) im Verwaltungshandbuch für Amazon EMR und [Spark-Programmierungs-Handbuch](#).
- [Hudi] Verbesserungen zur Vereinfachung der Hudi-Konfiguration. Optimistische Parallelitätssteuerung standardmäßig deaktiviert.

### Änderungen, Verbesserungen und behobene Probleme

- Dies ist eine Version zur Behebung von Problemen mit Amazon-EMR-Skalierung, wenn ein Cluster nicht erfolgreich hoch-/herunterskaliert werden kann oder Anwendungsfehler verursacht werden.
- Zuvor führte ein manueller Neustart des Ressourcenmanagers auf einem Multi-Master-Cluster dazu, dass Amazon-EMR-Cluster-Daemons wie Zookeeper alle zuvor stillgelegten oder verlorenen Knoten in der Zookeeper znode-Datei neu geladen haben. Dies führte dazu, dass die Standardgrenzwerte in bestimmten Situationen überschritten wurden. Amazon EMR entfernt jetzt



die außer Betrieb genommenen oder verlorenen Knotendatensätze, die älter als eine Stunde sind, aus der Zookeeper-Datei, und die internen Grenzwerte wurden erhöht.

- Es wurde ein Problem behoben, bei dem Skalierungsanforderungen für einen großen, stark ausgelasteten Cluster fehlschlagen, wenn Amazon-EMR-On-Cluster-Daemons Aktivitäten zur Integritätsprüfung durchführten, z. B. das Erfassen des YARN-Knotenstatus und des HDFS-Knotenstatus. Das lag daran, dass Cluster-Daemons die Integritätsstatusdaten eines Knotens nicht an interne Amazon-EMR-Komponenten weitergeben konnten.
- Verbesserte EMR-On-Cluster-Daemons zur korrekten Nachverfolgung der Knotenstatus bei der Wiederverwendung von IP-Adressen, um die Zuverlässigkeit bei Skalierungsvorgängen zu verbessern.
- [SPARK-29683](#). Es wurde ein Problem behoben, bei dem während der Cluster-Skalierung Auftragsfehler auftraten, da Spark davon ausging, dass alle verfügbaren Knoten auf der Verweigern-Liste standen.
- [YARN-9011](#). Es wurde ein Problem behoben, bei dem Auftragsfehler aufgrund eines Fehlers bei der Außerbetriebnahme von YARN auftraten, wenn der Cluster versuchte, hoch- oder herunterzuskalieren.
- Das Problem mit Schritt- oder Auftragsfehlern bei der Cluster-Skalierung wurde behoben, indem sichergestellt wurde, dass die Knotenstatus zwischen den Amazon-EMR-On-Cluster-Daemons und YARN/HDFS immer konsistent sind.
- Es wurde ein Problem behoben, bei dem Clustervorgänge wie Herunterskalierung und Schrittübermittlung für Amazon-EMR-Cluster, die mit Kerberos-Authentifizierung aktiviert waren, fehlschlagen. Dies lag daran, dass der Amazon-EMR-On-Cluster-Daemon das Kerberos-Ticket nicht erneuert hat, das für die sichere Kommunikation mit HDFS/YARN erforderlich ist, das auf dem Primärknoten ausgeführt wird.
- Zeppelin auf Version 0.10.0 upgegradet.
- Livy Fix – auf 0.7.1 aktualisiert
- Spark-Leistungsverbesserung – Heterogene Executors werden deaktiviert, wenn bestimmte Spark-Konfigurationswerte in EMR 5.34.0 überschrieben werden.
- WebHDFS- und HttpFS-Server werden standardmäßig deaktiviert. Sie können WebHDFS mithilfe der Hadoop-Konfiguration `dfs.webhdfs.enabled` erneut aktivieren. Der HttpFS-Server kann mit `sudo systemctl start hadoop-httpfs` gestartet werden.

## Bekannte Probleme

- Das Amazon-EMR-Notebooks-Feature, das mit dem Livy-Benutzerwechsel verwendet wird, funktioniert nicht, da HttpFS standardmäßig deaktiviert ist. In diesem Fall kann das EMR Notebook keine Verbindung zu dem Cluster herstellen, für den Livy-Identitätswechsel aktiviert ist. Die Problemumgehung besteht darin, den HttpFS-Server zu starten, bevor Sie das EMR-Notebook mit dem Cluster `sudo systemctl start hadoop-httpfs` verbinden.
- Hue-Abfragen funktionieren in Amazon EMR 6.4.0 nicht, da der Apache-Hadoop-HttpFS-Server standardmäßig deaktiviert ist. Um Hue auf Amazon EMR 6.4.0 zu verwenden, starten Sie den HttpFS-Server auf dem Amazon-EMR-Primärknoten entweder manuell mithilfe von `sudo systemctl start hadoop-httpfs` oder [verwenden Sie einen Amazon-EMR-Schritt](#).
- Das Amazon-EMR-Notebooks-Feature, das mit dem Livy-Benutzerwechsel verwendet wird, funktioniert nicht, da HttpFS standardmäßig deaktiviert ist. In diesem Fall kann das EMR Notebook keine Verbindung zu dem Cluster herstellen, für den Livy-Identitätswechsel aktiviert ist. Die Problemumgehung besteht darin, den HttpFS-Server zu starten, bevor Sie das EMR-Notebook mit dem Cluster `sudo systemctl start hadoop-httpfs` verbinden.
- Wenn Sie Spark mit der Formatierung des Hive-Partitionsstandorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den Amazon-EMR-Versionen 5.30.0 bis 5.36.0 und 6.2.0 bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:
  - Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.
  - Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist `s3://bucket/table/p=a` ein Präfix von `s3://bucket/table/p=a b`.
  - Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen UTF-8-Wert, der kleiner als das `/`-Zeichen (U+002F) ist. Beispielsweise fällt das Leerzeichen (U+0020), das in `s3://bucket/table/p=a b` zwischen `a` und `b` vorkommt, in diese Kategorie. Beachten Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: `!"#$%&'()*+,-.` Weitere Informationen finden Sie unter [UTF-8-Kodierungstabelle und Unicode-Zeichen](#).

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.

## 5.34.0 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-*amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.4.1	Amazon SageMaker Spark-SDK
<code>emr-ddb</code>	4.16.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.14.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.5.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-notebook-env</code>	1.4.0	Conda Env für EMR-Notebooks, das Jupyter Enterprise Gateway enthält
<code>emr-s3-dist-cp</code>	2.18.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.

Komponente	Version	Beschreibung
emr-s3-select	1.7.0	EMR S3Select-Konnektor
emrfs	2,48,0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.13.1	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
flink-jobmanager-config	1.13.1	Verwaltung von Ressourcen auf EMR-Knoten für Apache JobManager Flink.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.10.1-amzn-2	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.10.1-amzn-2	HDFS-Service auf Knotenebene zum Speichern von Blöcken.

Komponente	Version	Beschreibung
hadoop-hdfs-library	2.10.1-amzn-2	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.10.1-amzn-2	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-hdfs-journalnode	2.10.1-amzn-2	HDFS-Service zum Verwalten des Hadoop-Dateisystem journals auf HA-Clustern.
hadoop-httpfs-server	2.10.1-amzn-2	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.10.1-amzn-2	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.10.1-amzn-2	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.10.1-amzn-2	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.10.1-amzn-2	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.

Komponente	Version	Beschreibung
hadoop-yarn-timeline-server	2.10.1-amzn-2	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.4.13	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.4.13	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.4.13	HBase-Befehlszeilen-Client.
hbase-rest-server	1.4.13	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.4.13	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.8-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.8-amzn-0	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.3.8-amzn-0	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.

Komponente	Version	Beschreibung
hive-client	2.3.8-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.3.8-amzn-0	Hive-hbase client.
hive-metastore-server	2.3.8-amzn-0	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.3.8-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.9.0-amzn-0	Inkrementelles Verarbeitungs-Framework zur Versorgung der Datenpipeline mit geringer Latenz und hoher Effizienz.
hudi-spark	0.9.0-amzn-0	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.
hudi-presto	0.9.0-amzn-0	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hue-server	4.9.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	1.4.1	Multi-User-Server für Jupyter-Notebooks
livy-server	0.7.1-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark

Komponente	Version	Beschreibung
nginx	1.13.1	nginx [engine x] ist ein HTTP- und Reverse-Proxy-Server.
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.8.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.68	MySQL-Datenbankserver.
nvidia-cuda	10,1243	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.2.1	Oozie-Befehlszeilen-Client.
oozie-server	5.2.1	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	4.14.3-HBase-1.4	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.14.3-HBase-1.4	Ein schlanker Server für den Zugriff auf JDBC und Protokoll puffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.261-amzn-0	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.



Komponente	Version	Beschreibung
presto-worker	0.261-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.261-amzn-0	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	4.0.2	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	1.2.0	Apache Ranger Key Management System
spark-client	2.4.8-amzn-0	Spark-Befehlszeilen-Clients.
spark-history-server	2.4.8-amzn-0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.4.8-amzn-0	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.4.8-amzn-0	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.

Komponente	Version	Beschreibung
tensorflow	2.4.1	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.2	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.10.0	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.14	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.14	ZooKeeper Befehlszeilen-Cli-ent.

### 5.34.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Aktionen zur Neukonfiguration treten auf, wenn Sie eine Konfiguration für Instance-Gruppen in einem laufenden Cluster angeben. Amazon EMR initiiert nur Rekonfigurationsaktionen für die Klassifizierungen, die Sie ändern. Weitere Informationen finden Sie unter [Eine Instance-Gruppe in einem laufenden Cluster neu konfigurieren](#).

## emr-5.34.0-Klassifizierungen

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
capacity-scheduler	Ändert die Werte in der capacity-scheduler.xml-Datei in Hadoop.	Restarts the ResourceManager service.
container-executor	Ändern Sie die Werte in der Datei „container-executor.cfg“ Datei von Hadoop YARN.	Not available.
container-log4j	Ändert die Werte in der container-log4j.properties-Datei in Hadoop YARN.	Not available.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Ändern Sie die Docker-bezogenen Einstellungen.	Not available.
emrfs-site	Ändert die EMRFS-Einstellungen.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
		Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegionserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.	Restarts Flink history server.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.	Restarts Flink history server.
flink-log4j-session	Ändern Sie die Einstellungen von Flink log4j-session.properties für Kubernetes-/Yarn-Sitzungen.	Not available.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.	Restarts Flink history server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.	Not available.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Ändert die Werte in der HBase-Umgebung.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.	Not available.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.	Should not be reconfigured.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei proto-hive-site.xml von HCatalog.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.	Restarts Hive WebHCat Server.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.	Restarts Hive WebHCat Server.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.	Restarts Hive WebHCat Server.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.	Not available.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.	Not available.
hive-env	Ändert die Werte in der Hive-Umgebung.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.	Restarts HiveServer2 and HiveMetastore.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log	Not available.
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.	Not available.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.	Not available.
hue-ini	Ändert die Werte in der INI-Datei in Hue.	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.	Restarts Hadoop Httpfs service.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.	Not available.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.	Restarts Hadoop-KMS service.



Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.	Not available.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Ändern der Werte in der Hudi-Umgebung.	Not available.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.	Not available.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.	Not available.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.	Not available.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.	Not available.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.	Restarts Livy Server.
livy-env	Ändert die Werte in der Livy-Umgebung.	Restarts Livy Server.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.	Restarts Livy Server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Ändert die Werte in der Oozie-Umgebung.	Restarts Oozie.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.	Restarts Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.	Not available.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.	Not available.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.	Not available.
pig-env	Ändert die Werte in der Pig-Umgebung.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.	Restarts Oozie.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.	Not available.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.	Restarts Presto-Server.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.	Restarts Presto-Server.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.	Not available.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.	Restarts Presto-Server.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.	Not available.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.	Not available.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.	Not available.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.	Restarts Presto-Server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.	Not available.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.	Not available.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.	Not available.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.	Not available.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.	Not available.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.	Not available.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.	Not available.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.	Not available.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.	Not available.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.	Not available.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.	Not available.
ranger-kms-dbks-site	Ändert die Werte in der dbks-site.xml-Datei von Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Ändert die Werte in der Ranger KMS-Umgebung.	Restarts Ranger KMS Server.
ranger-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei von Ranger KMS.	Not available.
ranger-kms-db-ca	Ändert die Werte für die CA-Datei auf S3 für die MySQL SSL-Verbindung mit Ranger KMS.	Not available.
recordserver-env	Werte in der RecordServer EMR-Umgebung ändern.	Restarts EMR record server.
recordserver-conf	Ändern Sie die Werte in der Datei RecordServer server.properties von EMR.	Restarts EMR record server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
recordserver-log4j	Ändern Sie die Werte in der Datei RecordServer log4j.properties von EMR.	Restarts EMR record server.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Ändert die Werte in der Spark-Umgebung.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.	Not available.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.	Not available.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml von Sqoop OraOop.	Not available.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.	Restarts Oozie.
yarn-env	Ändert die Werte in der YARN-Umgebung.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.	Restarts Zeppelin.
zeppelin-site	Ändern Sie die Konfigurationseinstellungen in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.	Restarts Zookeeper server.

## Amazon-EMR-Version 5.33.1

### 5.33.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

#### Informationen zur Anwendungsversion

	emr-5.33.1	emr-5.33.0	emr-5.32.1	emr-5.32.0
AWS SDK for Java	1.11.970	1,11,970	1,11,890	1,11,890
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.12.1	1.12.1	1.11.2	1.11.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2



	emr-5.33.1	emr-5.33.0	emr-5.32.1	emr-5.32.0
HBase	1.4.13	1.4.13	1.4.13	1.4.13
HCatalog	2.3.7	2.3.7	2.3.7	2.3.7
Hadoop	2.10.1	2.10.1	2.10.1	2.10.1
Hive	2.3.7	2.3.7	2.3.7	2.3.7
Hudi	0.7.0-amzn-1	0.7.0-amzn-1	0.6.0-amzn-0	0.6.0-amzn-0
Hue	4.9.0	4.9.0	4.8.0	4.8.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	2.1.0	2.1.0	2.1.0	2.1.0
JupyterHub	1.2.2	1.2.2	1.1.0	1.1.0
Livy	0.7.0	0.7.0	0.7.0	0.7.0
MXNet	1.7.0	1.7.0	1.7.0	1.7.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.0	5.2.0	5.2.0	5.2.0
Phoenix	4.14.3	4.14.3	4.14.3	4.14.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0,245,1	0,245,1	0,240,1	0,240,1
Spark	2.4.7	2.4.7	2.4.7	2.4.7
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.4.1	2.4.1	2.3.1	2.3.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2

	emr-5.33.1	emr-5.33.0	emr-5.32.1	emr-5.32.0
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.9.0	0.9.0	0.8.2	0.8.2
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

### 5.33.1 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.33.0/5.33.1. Änderungen beziehen sich auf Version 5.32.0.

Erste Version: 19. April 2021

Letzte Aktualisierung: 9. August 2021

#### Upgrades

- Der Amazon-Glue-Konnektor wurde auf Version 1.15.0 aktualisiert
- Auf Version AWS SDK for Java 1.11.970 aktualisiert
- EMRFS auf Version 2.46.0. aktualisiert
- EMR Goodies auf Version 2.14.0 aktualisiert
- EMR Record Server auf Version 1.9.0 aktualisiert
- EMR S3 Dist CP auf Version 2.18.0 aktualisiert
- EMR Secret Agent Server auf Version 1.8.0 aktualisiert
- Flink auf Version 1.12.1 aktualisiert
- Hadoop wurde auf Version 2.10.1-amzn-1 aktualisiert
- Upgrade von Hive auf Version 2.3.7-amzn-4
- Hudi wurde auf Version 0.7.0 aktualisiert
- Hue auf Version 4.9.0 aktualisiert
- OpenCV wurde auf Version 4.5.0 aktualisiert
- Presto auf Version 0.245.1-amzn-0 aktualisiert
- R wurde auf Version 4.0.2 aktualisiert

- Spark auf Version 2.4.7-amzn-1 aktualisiert
- Auf Version 2.4.1 aktualisiert TensorFlow
- Upgrade von Zeppelin auf Version 0.9.0

## Änderungen, Verbesserungen und behobene Probleme

- Dies ist eine Version zur Behebung von Problemen mit Amazon-EMR-Skalierung, wenn ein Cluster nicht erfolgreich hoch-/herunterskaliert werden kann oder Anwendungsfehler verursacht werden.
- Es wurde ein Problem behoben, bei dem Skalierungsanforderungen für einen großen, stark ausgelasteten Cluster fehlschlugen, wenn Amazon-EMR-On-Cluster-Daemons Aktivitäten zur Integritätsprüfung durchführten, z. B. das Erfassen des YARN-Knotenstatus und des HDFS-Knotenstatus. Das lag daran, dass Cluster-Daemons die Integritätsstatusdaten eines Knotens nicht an interne Amazon-EMR-Komponenten weitergeben konnten.
- Verbesserte EMR-On-Cluster-Daemons zur korrekten Nachverfolgung der Knotenstatus bei der Wiederverwendung von IP-Adressen, um die Zuverlässigkeit bei Skalierungsvorgängen zu verbessern.
- [SPARK-29683](#). Es wurde ein Problem behoben, bei dem während der Cluster-Skalierung Auftragsfehler auftraten, da Spark davon ausging, dass alle verfügbaren Knoten auf der Verweigern-Liste standen.
- [YARN-9011](#). Es wurde ein Problem behoben, bei dem Auftragsfehler aufgrund eines Fehlers bei der Außerbetriebnahme von YARN auftraten, wenn der Cluster versuchte, hoch- oder herunterzuskalieren.
- Das Problem mit Schritt- oder Auftragsfehlern bei der Cluster-Skalierung wurde behoben, indem sichergestellt wurde, dass die Knotenstatus zwischen den Amazon-EMR-On-Cluster-Daemons und YARN/HDFS immer konsistent sind.
- Es wurde ein Problem behoben, bei dem Clustervorgänge wie Herunterskalierung und Schrittübermittlung für Amazon-EMR-Cluster, die mit Kerberos-Authentifizierung aktiviert waren, fehlschlugen. Dies lag daran, dass der Amazon-EMR-On-Cluster-Daemon das Kerberos-Ticket nicht erneuert hat, das für die sichere Kommunikation mit HDFS/YARN erforderlich ist, das auf dem Primärknoten ausgeführt wird.
- Neuere Amazon-EMR-Versionen beheben das Problem mit einem niedrigeren Limit für die „maximale Anzahl geöffneter Dateien“ auf älteren AL2-Versionen in Amazon EMR. Die Amazon-EMR-Versionen 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 und höher enthalten jetzt einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“.

- Konfigurieren eines Clusters zur Behebung von Leistungsproblemen mit Apache YARN Timeline Server Version 1 und 1.5

Apache YARN Timeline Server Version 1 und 1.5 können bei sehr aktiven, großen EMR-Clustern zu Leistungsproblemen führen, insbesondere bei `yarn.resourcemanager.system-metrics-publisher.enabled=true`, was die Standardeinstellung in Amazon EMR ist. Ein Open-Source-YARN-Timeline-Server v2 löst das Leistungsproblem im Zusammenhang mit der Skalierbarkeit des YARN Timeline Servers.

Weitere Lösungen für dieses Problem umfassen:

- Konfiguration von `yarn.resourcemanager.system-metrics-publisher.enabled=false` in `yarn-site.xml`.
- Aktivieren des Fixes für dieses Problem beim Erstellen eines Clusters, wie unten beschrieben.

Die folgenden Amazon-EMR-Versionen enthalten eine Lösung für dieses Leistungsproblem mit YARN Timeline Server.

EMR 5.30.2, 5.31.1, 5.32.1, 5.33.1, 5.34.x, 6.0.1, 6.1.1, 6.2.1, 6.3.1, 6.4.x

Um den Fix für eine der oben angegebenen Amazon-EMR-Versionen zu aktivieren, legen Sie diese Eigenschaften `true` in einer JSON-Konfigurationsdatei fest, die mit dem [aws emr create-cluster-Befehlsparameter](#) `--configurations file://./configurations.json` übergeben wird. Oder aktivieren Sie den Fix über die Benutzeroberfläche der [Rekonfigurationskonsole](#).

Beispiel für den Inhalt der Datei `configurations.json`:

```
[
{
  "Classification": "yarn-site",
  "Properties": {
    "yarn.resourcemanager.system-metrics-publisher.timeline-server-v1.enable-batch":
      "true",
    "yarn.resourcemanager.system-metrics-publisher.enabled": "true"
  },
  "Configurations": []
}
```

- Die Spark-Laufzeit ist jetzt schneller, wenn Partitionsspeicherorte aus Hive Metastore für Spark-Insert-Abfragen abgerufen werden.
- Aktualisierte Komponentenversionen. Eine Liste der Komponentenversionen finden Sie in diesem Handbuch unter [Über Amazon-EMR-Versionen](#).
- Das AWS Java SDK-Bundle wurde auf jedem neuen Cluster installiert. Dies ist ein einzelnes JAR, das alle Service-SDKs und ihre Abhängigkeiten enthält, anstatt einzelne Komponenten-Jars. Weitere Informationen finden Sie unter [Gebündelte Java SDK Abhängigkeit](#)
- Probleme bei der verwalteten Skalierung in früheren Amazon-EMR-Versionen wurden behoben und Verbesserungen vorgenommen, sodass die Ausfallraten von Anwendungen erheblich reduziert wurden.
- HTTPS ist jetzt standardmäßig für Amazon-Linux-Repositoryys aktiviert. Wenn Sie eine Amazon-S3-VPCE-Richtlinie verwenden, um den Zugriff auf bestimmte Buckets zu beschränken, müssen Sie den neuen Amazon-Linux-Bucket-ARN `arn:aws:s3:::amazonlinux-2-repos-$region/*` zu Ihrer Richtlinie hinzufügen (ersetzen Sie `$region` durch die Region, in der sich der Endpunkt befindet). Weitere Informationen finden Sie unter diesem Thema in den AWS Diskussionsforen. [Ankündigung: Amazon Linux 2 unterstützt jetzt die Möglichkeit, HTTPS zu verwenden, wenn eine Verbindung zu Paket-Repositoryys hergestellt wird.](#)

## Neue Features

- Amazon EMR unterstützt Amazon S3 Access Points, ein Feature von Amazon S3, mit dem Sie den Zugriff auf gemeinsam genutzte Data Lakes einfach verwalten können. Mit Ihrem Alias von Amazon S3 Access Point können Sie Ihren Datenzugriff auf Amazon EMR in großem Umfang vereinfachen. Sie können Amazon S3 Access Points mit allen Versionen von Amazon EMR ohne zusätzliche Kosten in allen AWS Regionen verwenden, in denen Amazon EMR verfügbar ist. Weitere Informationen zu Amazon S3 Access Points und Zugangspunkt-Aliassen finden Sie unter [Verwenden eines Alias im Bucket-Stil für Ihren Zugangspunkt](#) im Amazon-S3-Benutzerhandbuch.
- Amazon EMR-5.33 unterstützt neue Amazon-EC2-Instance-Typen: c5a, c5ad, c6gn, c6gd, m6gd, d3, d3en, m5zn, r5b, r6gd. Siehe [Unterstützte Instance-Typen](#).

## Bekannte Probleme

- Niedrigeres Limit für die „Max. Anzahl geöffneter Dateien“ auf älteren AL2-Versionen [in neueren Versionen behoben]. Amazon-EMR-Versionen: emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0 und emr-6.2.0 basieren auf älteren Versionen von Amazon Linux 2 (AL2), die eine

niedrigere Ulimit-Einstellung für „Max. Anzahl geöffneter Dateien“ haben, wenn Amazon-EMR-Cluster mit dem Standard-AMI erstellt werden. Die Amazon-EMR-Versionen 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 und höher enthalten einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“. Versionen mit einem niedrigeren Limit für geöffnete Dateien verursachen beim Senden des Spark-Jobs den Fehler „Zu viele offene Dateien“. In den betroffenen Versionen hat das Amazon-EMR-Standard-AMI eine Ulimit-StandardEinstellung von 4 096 für „Max. Anzahl geöffneter Dateien“, was unter dem Dateilimit von 65 536 im neuesten Amazon-Linux-2-AMI liegt. Die niedrigere Ulimit-Einstellung für „Max. Anzahl geöffneter Dateien“ führt dazu, dass der Spark-Job fehlschlägt, wenn der Spark-Treiber und der Executor versuchen, mehr als 4 096 Dateien zu öffnen. Um das Problem zu beheben, verfügt Amazon EMR über ein Bootstrap Action (BA)-Skript, das die Ulimit-Einstellung bei der Cluster-Erstellung anpasst.

Wenn Sie eine ältere Amazon-EMR-Version verwenden, für die dieses Problem nicht dauerhaft behoben ist, können Sie mit der folgenden Problemumgehung das Instance-Controller-Ulimit explizit auf maximal 65 536 Dateien festlegen.

Explizit ein ulimit über die Befehlszeile setzen

1. Bearbeiten Sie `/etc/systemd/system/instance-controller.service`, um die folgenden Parameter zum Abschnitt Service hinzuzufügen.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Starten Sie neu InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```


Mithilfe der Bootstrap-Aktion (BA) ein Ulimit festlegen

Sie können auch ein Bootstrap-Aktionsskript (BA) verwenden, um das Ulimit für den Instance-Controller bei der Clustererstellung auf 65 536 Dateien zu konfigurieren.

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
```

```
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

- Für private Subnetz-Cluster von Amazon EMR 6.3.0 und 6.2.0 können Sie nicht auf die Ganglia-Webbenutzeroberfläche zugreifen. Sie erhalten die Fehlermeldung „Zugriff verweigert (403)“. Andere Web-Benutzeroberflächen wie Spark, Hue JupyterHub, Zeppelin, Livy und Tez funktionieren normal. Der Zugriff auf die Ganglia-Web-Benutzeroberfläche auf öffentlichen Subnetzclustern funktioniert ebenfalls normal. Um dieses Problem zu beheben, starten Sie den httpd-Service auf dem Primärknoten mit `sudo systemctl restart httpd` neu. Dieses Problem wurde in Amazon EMR 6.4.0 behoben.

 **Important**

EMR-Cluster, auf denen Amazon-Linux- oder Amazon-Linux-2-AMIs (Amazon Machine Images) ausgeführt werden, verwenden das Standardverhalten von Amazon Linux und laden wichtige und kritische Kernel-Updates, die einen Neustart erfordern, nicht automatisch herunter und installieren sie. Dies ist dasselbe Verhalten wie bei anderen Amazon-EC2-Instances, die das standardmäßige Amazon-Linux-AMI ausführen. Wenn neue Amazon-Linux-Softwareupdates, die einen Neustart erfordern (wie Kernel-, NVIDIA- und CUDA-Updates), nach der Veröffentlichung einer Amazon-EMR-Version verfügbar werden, laden EMR-Cluster-Instances, die das Standard-AMI ausführen, diese Updates nicht automatisch herunter und installieren sie. Um Kernel-Updates zu erhalten, können Sie [Ihr Amazon-EMR-AMI](#) so anpassen, dass es [das neueste Amazon-Linux-AMI verwendet](#).

- Die Konsolenunterstützung zur Erstellung einer Sicherheitskonfiguration, die die AWS Ranger-Integrationsoption spezifiziert, wird derzeit in der Region nicht unterstützt. GovCloud Die Sicherheitskonfiguration kann mit der CLI durchgeführt werden. Weitere Informationen finden [Sie unter Erstellen der EMR-Sicherheitskonfiguration](#) im Verwaltungshandbuch für Amazon EMR.

- **Verwaltete Richtlinien mit Geltungsbereich:** Um den AWS bewährten Methoden zu entsprechen, hat Amazon EMR verwaltete Standardrichtlinien mit EMR-Geltungsbereich der Version v2 eingeführt, die als Ersatz für Richtlinien dienen, die nicht mehr unterstützt werden. Weitere Informationen finden Sie unter [Verwaltete Richtlinien von Amazon EMR](#).
- Wenn Sie Spark mit der Formatierung des Hive-Partitionsstandorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den Amazon-EMR-Versionen 5.30.0 bis 5.36.0 und 6.2.0 bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:
  - Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.
  - Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist `s3://bucket/table/p=a` ein Präfix von `s3://bucket/table/p=a b`.
  - Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen UTF-8-Wert, der kleiner als das /-Zeichen (U+002F) ist. Beispielsweise fällt das Leerzeichen (U+0020), das in `s3://bucket/table/p=a b` zwischen a und b vorkommt, in diese Kategorie. Beachten Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: `!"#$%&'()*+,-.` Weitere Informationen finden Sie unter [UTF-8-Kodierungstabelle und Unicode-Zeichen](#).

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.

### 5.33.1 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die



Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als 2.2-amzn-2 aufgeführt.

Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1.4.1	Amazon SageMaker Spark-SDK
emr-ddb	4.16.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.14.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.5.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-notebook-env	1.2.0	Conda Env für EMR-Notebooks, das Jupyter Enterprise Gateway enthält
emr-s3-dist-cp	2.18.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	1.6.0	EMR S3Select-Konnektor
emrfs	2,46,0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.12.1	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.

Komponente	Version	Beschreibung
flink-jobmanager-config	1.12.1	Verwaltung von Ressourcen auf EMR-Knoten für Apache JobManager Flink.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.10.1-amzn-1.1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.10.1-amzn-1.1	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.10.1-amzn-1.1	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.10.1-amzn-1.1	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.

Komponente	Version	Beschreibung
hadoop-hdfs-journalnode	2.10.1-amzn-1.1	HDFS-Service zum Verwalten des Hadoop-Dateisystem journals auf HA-Clustern.
hadoop-httpfs-server	2.10.1-amzn-1.1	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.10.1-amzn-1.1	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.10.1-amzn-1.1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.10.1-amzn-1.1	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.10.1-amzn-1.1	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.10.1-amzn-1.1	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.4.13	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.

Komponente	Version	Beschreibung
hbase-region-server	1.4.13	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.4.13	HBase-Befehlszeilen-Client.
hbase-rest-server	1.4.13	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.4.13	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.7-amzn-4	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.7-amzn-4	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.3.7-amzn-4	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.3.7-amzn-4	Hive-Befehlszeilen-Client.
hive-hbase	2.3.7-amzn-4	Hive-hbase client.
hive-metastore-server	2.3.7-amzn-4	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).

Komponente	Version	Beschreibung
hive-server2	2.3.7-amzn-4	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.7.0-amzn-1	Inkrementelles Verarbeitungs-Framework zur Stromversorgung der Datenpipeline bei geringer Latenz und hoher Effizienz.
hudi-spark	0.7.0-amzn-1	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.
hudi-presto	0.7.0-amzn-1	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hue-server	4.9.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	1.2.2	Multi-User-Server für Jupyter-Notebooks
livy-server	0.7.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP- und Reverse-Proxy-Server.
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.7.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.

Komponente	Version	Beschreibung
mariadb-server	5.5.68+	MySQL-Datenbankserver.
nvidia-cuda	10.1,243	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.2.0	Oozie-Befehlszeilen-Client.
oozie-server	5.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	4.14.3-HBase-1.4	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.14.3-HBase-1.4	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.245.1-amzn-0	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.245.1-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.245.1-amzn-0	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.

Komponente	Version	Beschreibung
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	4.0.2	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	1.2.0	Apache Ranger Key Management System
spark-client	2.4.7-amzn-1.1	Spark-Befehlszeilen-Clients.
spark-history-server	2.4.7-amzn-1.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.4.7-amzn-1.1	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.4.7-amzn-1.1	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	2.4.1	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.2	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.

Komponente	Version	Beschreibung
zeppelin-server	0.9.0	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.14	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.14	ZooKeeper Befehlszeilen-Cli-ent.

### 5.33.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Aktionen zur Neukonfiguration treten auf, wenn Sie eine Konfiguration für Instance-Gruppen in einem laufenden Cluster angeben. Amazon EMR initiiert nur Rekonfigurationsaktionen für die Klassifizierungen, die Sie ändern. Weitere Informationen finden Sie unter [Eine Instance-Gruppe in einem laufenden Cluster neu konfigurieren](#).

#### emr-5.33.1-Klassifizierungen

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.	Restarts the Resource Manager service.



Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
container-executor	Ändern Sie die Werte in der Datei „container-executor.cfg“ Datei von Hadoop YARN.	Not available.
container-log4j	Ändert die Werte in der container-log4j.properties-Datei in Hadoop YARN.	Not available.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Ändern Sie die Docker-bezogenen Einstellungen.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
emrfs-site	Ändert die EMRFS-Einstellungen.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegistrator, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.	Restarts Flink history server.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.	Restarts Flink history server.
flink-log4j-session	Ändern Sie die Einstellungen von Flink log4j-session.properties für Kubernetes-/Yarn-Sitzungen.	Not available.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.	Restarts Flink history server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.	Not available.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Ändert die Werte in der HBase-Umgebung.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.	Not available.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.	Should not be reconfigured.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei proto-hive-site.xml von HCatalog.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.	Restarts Hive WebHCat Server.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.	Restarts Hive WebHCat Server.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.	Restarts Hive WebHCat Server.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.	Not available.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.	Not available.
hive-env	Ändert die Werte in der Hive-Umgebung.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.	Restarts HiveServer2 and HiveMetastore.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log	Not available.
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.	Not available.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.	Not available.
hue-ini	Ändert die Werte in der INI-Datei in Hue.	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.	Restarts Hadoop Httpfs service.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.	Not available.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.	Restarts Hadoop-KMS service.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.	Not available.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Ändern der Werte in der Hudi-Umgebung.	Not available.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.	Not available.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.	Not available.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.	Not available.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.	Not available.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.	Restarts Livy Server.
livy-env	Ändert die Werte in der Livy-Umgebung.	Restarts Livy Server.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.	Restarts Livy Server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
mapred-env	Ändern Sie die Werte in der MapReduce Anwendung sumgebung.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Ändert die Werte in der Oozie-Umgebung.	Restarts Oozie.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.	Restarts Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.	Not available.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.	Not available.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.	Not available.
pig-env	Ändert die Werte in der Pig-Umgebung.	Not available.



Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.	Restarts Oozie.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.	Not available.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.	Restarts Presto-Server.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.	Restarts Presto-Server.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.	Not available.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.	Restarts Presto-Server.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.	Not available.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.	Not available.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.	Not available.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.	Restarts Presto-Server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.	Not available.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.	Not available.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.	Not available.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.	Not available.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.	Not available.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.	Not available.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.	Not available.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.	Not available.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.	Not available.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.	Not available.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.	Not available.
ranger-kms-dbks-site	Ändert die Werte in der dbks-site.xml-Datei von Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Ändert die Werte in der Ranger KMS-Umgebung.	Restarts Ranger KMS Server.
ranger-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei von Ranger KMS.	Not available.
ranger-kms-db-ca	Ändert die Werte für die CA-Datei auf S3 für die MySQL SSL-Verbindung mit Ranger KMS.	Not available.
recordserver-env	Werte in der RecordServer EMR-Umgebung ändern.	Restarts EMR record server.
recordserver-conf	Ändern Sie die Werte in der Datei RecordServer server.properties von EMR.	Restarts EMR record server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
recordserver-log4j	Ändern Sie die Werte in der Datei RecordServer log4j.properties von EMR.	Restarts EMR record server.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Ändert die Werte in der Spark-Umgebung.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.	Not available.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.	Not available.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml von Sqoop OraOop.	Not available.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.	Restarts Oozie.
yarn-env	Ändert die Werte in der YARN-Umgebung.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.	Restarts Zeppelin.
zeppelin-site	Ändern Sie die Konfigurationseinstellungen in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.	Restarts Zookeeper server.

## Amazon-EMR-Version 5.33.0

### 5.33.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

#### Informationen zur Anwendungsversion

	emr-5.33.0	emr-5.32.1	emr-5.32.0	emr-5.31.1
AWS SDK for Java	1.11.970	1,11,890	1,11,890	1.11.852
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.12.1	1.11.2	1.11.2	1.11.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2

	emr-5.33.0	emr-5.32.1	emr-5.32.0	emr-5.31.1
HBase	1.4.13	1.4.13	1.4.13	1.4.13
HCatalog	2.3.7	2.3.7	2.3.7	2.3.7
Hadoop	2.10.1	2.10.1	2.10.1	2.10.0
Hive	2.3.7	2.3.7	2.3.7	2.3.7
Hudi	0.7.0-amzn-1	0.6.0-amzn-0	0.6.0-amzn-0	0.6.0-amzn-0
Hue	4.9.0	4.8.0	4.8.0	4.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	2.1.0	2.1.0	2.1.0	-
JupyterHub	1.2.2	1.1.0	1.1.0	1.1.0
Livy	0.7.0	0.7.0	0.7.0	0.7.0
MXNet	1.7.0	1.7.0	1.7.0	1.6.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.0	5.2.0	5.2.0	5.2.0
Phoenix	4.14.3	4.14.3	4.14.3	4.14.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0,245,1	0,240,1	0,240,1	0,238,3
Spark	2.4.7	2.4.7	2.4.7	2.4.6
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.4.1	2.3.1	2.3.1	2.1.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2

	emr-5.33.0	emr-5.32.1	emr-5.32.0	emr-5.31.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.9.0	0.8.2	0.8.2	0.8.2
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

## 5.33.0 Versionshinweise

### 5.33.0 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.4.1	Amazon SageMaker Spark-SDK
<code>emr-ddb</code>	4.16.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.14.0	Praktische Bibliotheken für das Hadoop-Ökosystem.



Komponente	Version	Beschreibung
emr-kinesis	3.5.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-notebook-env	1.2.0	Conda Env für EMR-Notebooks, das Jupyter Enterprise Gateway enthält
emr-s3-dist-cp	2.18.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	1.6.0	EMR S3Select-Konnektor
emrfs	2,46,0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.12.1	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
flink-jobmanager-config	1.12.1	Verwaltung von Ressourcen auf EMR-Knoten für Apache JobManager Flink.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.

Komponente	Version	Beschreibung
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.10.1-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.10.1-amzn-1	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.10.1-amzn-1	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.10.1-amzn-1	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-hdfs-journalnode	2.10.1-amzn-1	HDFS-Service zum Verwalten des Hadoop-Dateisystem journals auf HA-Clustern.
hadoop-httpfs-server	2.10.1-amzn-1	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.10.1-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.10.1-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.

Komponente	Version	Beschreibung
hadoop-yarn-nodemanager	2.10.1-amzn-1	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.10.1-amzn-1	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.10.1-amzn-1	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.4.13	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.4.13	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.4.13	HBase-Befehlszeilen-Client.
hbase-rest-server	1.4.13	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.4.13	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.7-amzn-4	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.

Komponente	Version	Beschreibung
hcatalog-server	2.3.7-amzn-4	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.3.7-amzn-4	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.3.7-amzn-4	Hive-Befehlszeilen-Client.
hive-hbase	2.3.7-amzn-4	Hive-hbase client.
hive-metastore-server	2.3.7-amzn-4	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.3.7-amzn-4	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.7.0-amzn-1	Inkrementelles Verarbeitungs-Framework zur Stromversorgung der Datenpipeline bei geringer Latenz und hoher Effizienz.
hudi-spark	0.7.0-amzn-1	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.
hudi-presto	0.7.0-amzn-1	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.

Komponente	Version	Beschreibung
hue-server	4.9.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	1.2.2	Multi-User-Server für Jupyter-Notebooks
livy-server	0.7.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP- und Reverse-Proxy-Server.
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.7.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.68	MySQL-Datenbankserver.
nvidia-cuda	10,1243	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.2.0	Oozie-Befehlszeilen-Client.
oozie-server	5.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	4.5.0	Open Source Computer Vision Library.
phoenix-library	4.14.3-HBase-1.4	Die Phoenix-Bibliotheken für den Server und den Client

Komponente	Version	Beschreibung
phoenix-query-server	4.14.3-HBase-1.4	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.245.1-amzn-0	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.245.1-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.245.1-amzn-0	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	4.0.2	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	1.2.0	Apache Ranger Key Management System
spark-client	2.4.7-amzn-1	Spark-Befehlszeilen-Clients.
spark-history-server	2.4.7-amzn-1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.

Komponente	Version	Beschreibung
spark-on-yarn	2.4.7-amzn-1	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.4.7-amzn-1	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	2.4.1	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.2	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.9.0	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.14	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.14	ZooKeeper Befehlszeilen-Client.

## 5.33.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Aktionen zur Neukonfiguration treten auf, wenn Sie eine Konfiguration für Instance-Gruppen in einem laufenden Cluster angeben. Amazon EMR initiiert nur Rekonfigurationsaktionen für die Klassifizierungen, die Sie ändern. Weitere Informationen finden Sie unter [Eine Instance-Gruppe in einem laufenden Cluster neu konfigurieren](#).

### emr-5.33.0-Klassifizierungen

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.	Restarts the Resource Manager service.
container-executor	Ändern Sie die Werte in der Datei „ <code>container-executor.cfg</code> “ Datei von Hadoop YARN.	Not available.
container-log4j	Ändert die Werte in der <code>container-log4j.properties</code> -Datei in Hadoop YARN.	Not available.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop



Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
		Httpfs, and MapReduce-HistoryServer.
docker-conf	Ändern Sie die Docker-bezogenen Einstellungen.	Not available.
emrfs-site	Ändert die EMRFS-Einstellungen.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegistrator, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.	Restarts Flink history server.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.	Restarts Flink history server.
flink-log4j-session	Ändern Sie die Einstellungen von Flink log4j-session.properties für Kubernetes-/Yarn-Sitzungen.	Not available.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.	Restarts Flink history server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.	Not available.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Ändert die Werte in der HBase-Umgebung.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.	Not available.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.	Should not be reconfigured.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei proto-hive-site .xml von HCatalog.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.	Restarts Hive WebHCat Server.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.	Restarts Hive WebHCat Server.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.	Restarts Hive WebHCat Server.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.	Not available.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.	Not available.
hive-env	Ändert die Werte in der Hive-Umgebung.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.	Restarts HiveServer2 and HiveMetastore.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log	Not available.
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.	Not available.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.	Not available.
hue-ini	Ändert die Werte in der INI-Datei in Hue.	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.	Restarts Hadoop Httpfs service.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.	Not available.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.	Restarts Hadoop-KMS service.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.	Not available.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Ändern der Werte in der Hudi-Umgebung.	Not available.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.	Not available.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.	Not available.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.	Not available.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.	Not available.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.	Restarts Livy Server.
livy-env	Ändert die Werte in der Livy-Umgebung.	Restarts Livy Server.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.	Restarts Livy Server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
mapred-env	Ändern Sie die Werte in der MapReduce Anwendung sumgebung.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Ändert die Werte in der Oozie-Umgebung.	Restarts Oozie.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.	Restarts Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.	Not available.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.	Not available.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.	Not available.
pig-env	Ändert die Werte in der Pig-Umgebung.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.	Restarts Oozie.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.	Not available.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.	Restarts Presto-Server.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.	Restarts Presto-Server.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.	Not available.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.	Restarts Presto-Server.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.	Not available.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.	Not available.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.	Not available.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.	Restarts Presto-Server.



Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.	Not available.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.	Not available.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.	Not available.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.	Not available.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.	Not available.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.	Not available.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.	Not available.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.	Not available.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.	Not available.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.	Not available.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.	Not available.
ranger-kms-dbks-site	Ändert die Werte in der dbks-site.xml-Datei von Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Ändert die Werte in der Ranger KMS-Umgebung.	Restarts Ranger KMS Server.
ranger-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei von Ranger KMS.	Not available.
ranger-kms-db-ca	Ändert die Werte für die CA-Datei auf S3 für die MySQL SSL-Verbindung mit Ranger KMS.	Not available.
recordserver-env	Werte in der RecordServer EMR-Umgebung ändern.	Restarts EMR record server.
recordserver-conf	Ändern Sie die Werte in der Datei RecordServer server.properties von EMR.	Restarts EMR record server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
recordserver-log4j	Ändern Sie die Werte in der Datei RecordServer log4j.properties von EMR.	Restarts EMR record server.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Ändert die Werte in der Spark-Umgebung.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.	Not available.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.	Not available.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml von Sqoop OraOop.	Not available.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.	Restarts Oozie.
yarn-env	Ändert die Werte in der YARN-Umgebung.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.	Restarts Zeppelin.
zeppelin-site	Ändern Sie die Konfigurationseinstellungen in zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.	Restarts Zookeeper server.

## Amazon-EMR-Version 5.32.1

### 5.32.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

#### Informationen zur Anwendungsversion

	emr-5.32.1	emr-5.32.0	emr-5.31.1	emr-5.31.0
AWS SDK for Java	1.11.890	1,11,890	1.11.852	1,11.852
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.11.2	1.11.2	1.11.0	1.11.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2

	emr-5.32.1	emr-5.32.0	emr-5.31.1	emr-5.31.0
HBase	1.4.13	1.4.13	1.4.13	1.4.13
HCatalog	2.3.7	2.3.7	2.3.7	2.3.7
Hadoop	2.10.1	2.10.1	2.10.0	2.10.0
Hive	2.3.7	2.3.7	2.3.7	2.3.7
Hudi	0.6.0-amzn-0	0.6.0-amzn-0	0.6.0-amzn-0	0.6.0-amzn-0
Hue	4.8.0	4.8.0	4.7.1	4.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	2.1.0	2.1.0	-	-
JupyterHub	1.1.0	1.1.0	1.1.0	1.1.0
Livy	0.7.0	0.7.0	0.7.0	0.7.0
MXNet	1.7.0	1.7.0	1.6.0	1.6.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.0	5.2.0	5.2.0	5.2.0
Phoenix	4.14.3	4.14.3	4.14.3	4.14.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0,240,1	0,240,1	0,238,3	0,238,3
Spark	2.4.7	2.4.7	2.4.6	2.4.6
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.3.1	2.3.1	2.1.0	2.1.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2

	emr-5.32.1	emr-5.32.0	emr-5.31.1	emr-5.31.0
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.2	0.8.2	0.8.2	0.8.2
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

### 5.32.1 Versionshinweise

Dies ist eine Version zur Behebung von Problemen mit Amazon-EMR-Skalierung, wenn ein Cluster nicht erfolgreich hoch-/herunterskaliert werden kann oder Anwendungsfehler verursacht werden.

#### Änderungen, Verbesserungen und behobene Probleme

- Es wurde ein Problem behoben, bei dem Skalierungsanforderungen für einen großen, stark ausgelasteten Cluster fehlschlagen, wenn Amazon-EMR-On-Cluster-Daemons Aktivitäten zur Integritätsprüfung durchführten, z. B. das Erfassen des YARN-Knotenstatus und des HDFS-Knotenstatus. Das lag daran, dass Cluster-Daemons die Integritätsstatusdaten eines Knotens nicht an interne Amazon-EMR-Komponenten weitergeben konnten.
- Verbesserte EMR-On-Cluster-Daemons zur korrekten Nachverfolgung der Knotenstatus bei der Wiederverwendung von IP-Adressen, um die Zuverlässigkeit bei Skalierungsvorgängen zu verbessern.
- [SPARK-29683](#). Es wurde ein Problem behoben, bei dem während der Cluster-Skalierung Auftragsfehler auftraten, da Spark davon ausging, dass alle verfügbaren Knoten auf der Verweigern-Liste standen.
- [YARN-9011](#). Es wurde ein Problem behoben, bei dem Auftragsfehler aufgrund eines Fehlers bei der Außerbetriebnahme von YARN auftraten, wenn der Cluster versuchte, hoch- oder herunterzuskalieren.
- Das Problem mit Schritt- oder Auftragsfehlern bei der Cluster-Skalierung wurde behoben, indem sichergestellt wurde, dass die Knotenstatus zwischen den Amazon-EMR-On-Cluster-Daemons und YARN/HDFS immer konsistent sind.
- Es wurde ein Problem behoben, bei dem Clustervorgänge wie Herunterskalierung und Schrittübermittlung für Amazon-EMR-Cluster, die mit Kerberos-Authentifizierung aktiviert waren, fehlschlagen. Dies lag daran, dass der Amazon-EMR-On-Cluster-Daemon das Kerberos-Ticket

nicht erneuert hat, das für die sichere Kommunikation mit HDFS/YARN erforderlich ist, das auf dem Primärknoten ausgeführt wird.

- Neuere Amazon-EMR-Versionen beheben das Problem mit einem niedrigeren Limit für die „maximale Anzahl geöffneter Dateien“ auf älteren AL2-Versionen in Amazon EMR. Die Amazon-EMR-Versionen 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 und höher enthalten jetzt einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“.
- HTTPS ist jetzt standardmäßig für Amazon-Linux-Repositorys aktiviert. Wenn Sie eine Amazon-S3-VPCE-Richtlinie verwenden, um den Zugriff auf bestimmte Buckets zu beschränken, müssen Sie den neuen Amazon-Linux-Bucket-ARN `arn:aws:s3:::amazonlinux-2-repos-$region/*` zu Ihrer Richtlinie hinzufügen (ersetzen Sie `$region` durch die Region, in der sich der Endpunkt befindet). Weitere Informationen finden Sie unter diesem Thema in den AWS Diskussionsforen. [Ankündigung: Amazon Linux 2 unterstützt jetzt die Möglichkeit, HTTPS zu verwenden, wenn eine Verbindung zu Paket-Repositorys hergestellt wird.](#)

## Bekannte Probleme

- Wenn Sie Spark mit der Formatierung des Hive-Partitionsstandorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den Amazon-EMR-Versionen 5.30.0 bis 5.36.0 und 6.2.0 bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:
  - Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.
  - Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist `s3://bucket/table/p=a` ein Präfix von `s3://bucket/table/p=a b`.
  - Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen UTF-8-Wert, der kleiner als das `/`-Zeichen (U+002F) ist. Beispielsweise fällt das Leerzeichen (U+0020), das in `s3://bucket/table/p=a b` zwischen `a` und `b` vorkommt, in diese Kategorie. Beachten Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: `!"#$%&'()*+,-.` Weitere Informationen finden Sie unter [UTF-8-Kodierungstabelle und Unicode-Zeichen](#).

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.



### 5.32.1 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.4.1	Amazon SageMaker Spark-SDK
<code>emr-ddb</code>	4.16.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.13.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.5.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-notebook-env</code>	1.1.0	Conda Env für EMR-Notebooks, das Jupyter Enterprise Gateway enthält
<code>emr-s3-dist-cp</code>	2.17.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.

Komponente	Version	Beschreibung
emr-s3-select	1.6.0	EMR S3Select-Konnektor
emrfs	2,45,0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.11.2	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
flink-jobmanager-config	1.11.2	Verwaltung von Ressourcen auf EMR-Knoten für Apache JobManager Flink.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.10.1-amzn-0.1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.10.1-amzn-0.1	HDFS-Service auf Knotenebene zum Speichern von Blöcken.

Komponente	Version	Beschreibung
hadoop-hdfs-library	2.10.1-amzn-0.1	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.10.1-amzn-0.1	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-hdfs-journalnode	2.10.1-amzn-0.1	HDFS-Service zum Verwalten des Hadoop-Dateisystem journals auf HA-Clustern.
hadoop-httfs-server	2.10.1-amzn-0.1	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.10.1-amzn-0.1	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.10.1-amzn-0.1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.10.1-amzn-0.1	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.10.1-amzn-0.1	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.

Komponente	Version	Beschreibung
hadoop-yarn-timeline-server	2.10.1-amzn-0.1	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.4.13	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.4.13	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.4.13	HBase-Befehlszeilen-Client.
hbase-rest-server	1.4.13	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.4.13	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.7-amzn-3	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.7-amzn-3	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.3.7-amzn-3	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.

Komponente	Version	Beschreibung
hive-client	2.3.7-amzn-3	Hive-Befehlszeilen-Client.
hive-hbase	2.3.7-amzn-3	Hive-hbase client.
hive-metastore-server	2.3.7-amzn-3	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.3.7-amzn-3	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.6.0-amzn-0	Inkrementelles Verarbeitungs-Framework zur Stromversorgung der Datenpipeline bei geringer Latenz und hoher Effizienz.
hudi-spark	0.6.0-amzn-0	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.
hudi-presto	0.6.0-amzn-0	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hue-server	4.8.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	1.1.0	Multi-User-Server für Jupyter-Notebooks
livy-server	0.7.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark

Komponente	Version	Beschreibung
nginx	1.12.1	nginx [engine x] ist ein HTTP- und Reverse-Proxy-Server.
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.7.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.68+	MySQL-Datenbankserver.
nvidia-cuda	10.1,243	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.2.0	Oozie-Befehlszeilen-Client.
oozie-server	5.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	4.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.3-HBase-1.4	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.14.3-HBase-1.4	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.240.1-amzn-0	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.

Komponente	Version	Beschreibung
presto-worker	0.240.1-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.240.1-amzn-0	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.3	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	1.2.0	Apache Ranger Key Management System
spark-client	2.4.7-amzn-0.1	Spark-Befehlszeilen-Clients.
spark-history-server	2.4.7-amzn-0.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.4.7-amzn-0.1	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.4.7-amzn-0.1	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.

Komponente	Version	Beschreibung
tensorflow	2.3.1	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.2	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.8.2	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.14	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.14	ZooKeeper Befehlszeilen-Cli-ent.

### 5.32.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Aktionen zur Neukonfiguration treten auf, wenn Sie eine Konfiguration für Instance-Gruppen in einem laufenden Cluster angeben. Amazon EMR initiiert nur Rekonfigurationsaktionen für die Klassifizierungen, die Sie ändern. Weitere Informationen finden Sie unter [Eine Instance-Gruppe in einem laufenden Cluster neu konfigurieren](#).



## emr-5.32.1-Klassifizierungen

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
capacity-scheduler	Ändert die Werte in der capacity-scheduler.xml-Datei in Hadoop.	Restarts the ResourceManager service.
container-executor	Ändern Sie die Werte in der Datei „container-executor.cfg“ Datei von Hadoop YARN.	Not available.
container-log4j	Ändert die Werte in der container-log4j.properties-Datei in Hadoop YARN.	Not available.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Ändern Sie die Docker-bezogenen Einstellungen.	Not available.
emrfs-site	Ändert die EMRFS-Einstellungen.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
		Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegistrator, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.	Restarts Flink history server.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.	Restarts Flink history server.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4j-yarn-session.properties.	Not available.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.	Restarts Flink history server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.	Not available.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Ändert die Werte in der HBase-Umgebung.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.	Not available.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.	Should not be reconfigured.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.	Restarts Hive WebHCat Server.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.	Restarts Hive WebHCat Server.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.	Restarts Hive WebHCat Server.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.	Not available.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.	Not available.
hive-env	Ändert die Werte in der Hive-Umgebung.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.	Restarts HiveServer2 and HiveMetastore.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log	Not available.
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.	Not available.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.	Not available.
hue-ini	Ändert die Werte in der INI-Datei in Hue.	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.	Restarts Hadoop Httpfs service.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.	Not available.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.	Restarts Hadoop-KMS service.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.	Not available.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Ändern der Werte in der Hudi-Umgebung.	Not available.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.	Not available.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.	Not available.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.	Not available.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.	Not available.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.	Restarts Livy Server.
livy-env	Ändert die Werte in der Livy-Umgebung.	Restarts Livy Server.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.	Restarts Livy Server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
mapred-env	Ändern Sie die Werte in der MapReduce Anwendung sumgebung.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Ändert die Werte in der Oozie-Umgebung.	Restarts Oozie.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.	Restarts Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.	Not available.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.	Not available.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.	Not available.
pig-env	Ändert die Werte in der Pig-Umgebung.	Not available.



Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.	Restarts Oozie.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.	Not available.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.	Restarts Presto-Server.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.	Restarts Presto-Server.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.	Not available.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.	Restarts Presto-Server.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.	Not available.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.	Not available.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.	Not available.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.	Restarts Presto-Server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.	Not available.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.	Not available.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.	Not available.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.	Not available.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.	Not available.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.	Not available.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.	Not available.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.	Not available.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.	Not available.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.	Not available.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.	Not available.
ranger-kms-dbks-site	Ändert die Werte in der dbks-site.xml-Datei von Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Ändert die Werte in der Ranger KMS-Umgebung.	Restarts Ranger KMS Server.
ranger-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei von Ranger KMS.	Not available.
ranger-kms-db-ca	Ändert die Werte für die CA-Datei auf S3 für die MySQL SSL-Verbindung mit Ranger KMS.	Not available.
recordserver-env	Werte in der RecordServer EMR-Umgebung ändern.	Restarts EMR record server.
recordserver-conf	Ändern Sie die Werte in der Datei RecordServer server.properties von EMR.	Restarts EMR record server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
recordserver-log4j	Ändern Sie die Werte in der Datei RecordServer log4j.properties von EMR.	Restarts EMR record server.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Ändert die Werte in der Spark-Umgebung.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.	Not available.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.	Not available.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml von Sqoop OraOop.	Not available.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.	Restarts Oozie.
yarn-env	Ändert die Werte in der YARN-Umgebung.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.	Restarts Zeppelin.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.	Restarts Zookeeper server.

## Amazon-EMR-Version 5.32.0

### 5.32.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

#### Informationen zur Anwendungsversion

	emr-5.32.0	emr-5.31.1	emr-5.31.0	emr-5.30.2
AWS SDK for Java	1.11.890	1.11.852	1,11.852	1,11.759
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.11.2	1.11.0	1.11.0	1.10.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.13	1.4.13	1.4.13	1.4.13
HCatalog	2.3.7	2.3.7	2.3.7	2.3.6
Hadoop	2.10.1	2.10.0	2.10.0	2.8.5
Hive	2.3.7	2.3.7	2.3.7	2.3.6

	emr-5.32.0	emr-5.31.1	emr-5.31.0	emr-5.30.2
Hudi	0.6.0-amzn-0	0.6.0-amzn-0	0.6.0-amzn-0	0.5.2-incubating
Hue	4.8.0	4.7.1	4.7.1	4.6.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	2.1.0	-	-	-
JupyterHub	1.1.0	1.1.0	1.1.0	1.1.0
Livy	0.7.0	0.7.0	0.7.0	0.7.0
MXNet	1.7.0	1.6.0	1.6.0	1.5.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.0	5.2.0	5.2.0	5.2.0
Phoenix	4.14.3	4.14.3	4.14.3	4.14.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0,240,1	0,238,3	0,238,3	0.232
Spark	2.4.7	2.4.6	2.4.6	2.4.5
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.3.1	2.1.0	2.1.0	1.14.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.2	0.8.2	0.8.2	0.8.2
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

## 5.32.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.32.0. Änderungen beziehen sich auf Version 5.31.0.

Erste Version: 8. Januar 2021

### Upgrades

- Der Amazon-Glue-Konnektor wurde auf Version 1.14.0 aktualisiert
- Amazon SageMaker Spark SDK wurde auf Version 1.4.1 aktualisiert
- Auf Version AWS SDK for Java 1.11.890 aktualisiert
- EMR DynamoDB Connector wurde auf Version 4.16.0 aktualisiert
- EMRFS auf Version 2.45.0 aktualisiert
- EMR-Log-Analytics-Metriken wurden auf Version 1.18.0 aktualisiert
- MetricsAndEventsApiGateway EMR-Client wurde auf Version 1.5.0 aktualisiert
- EMR Record Server auf Version 1.8.0 aktualisiert
- EMR S3 Dist CP auf Version 2.17.0 aktualisiert
- EMR Record Server auf Version 1.7.0 aktualisiert
- Flink auf Version 1.11.2 aktualisiert
- Hadoop wurde auf Version 2.10.1-amzn-0 aktualisiert
- Upgrade von Hive auf Version 2.3.7-amzn-3
- Hue auf Version 4.8.0 aktualisiert
- Mxnet wurde auf Version 1.7.0 aktualisiert
- OpenCV wurde auf Version 4.4.0 aktualisiert
- Presto auf Version 0.240.1-amzn-0 aktualisiert
- Spark auf Version 2.4.7-amzn-0 aktualisiert
- Auf Version TensorFlow 2.3.1 aktualisiert

### Änderungen, Verbesserungen und behobene Probleme

- Dies ist eine Version zur Behebung von Problemen mit Amazon-EMR-Skalierung, wenn ein Cluster nicht erfolgreich hoch-/herunterskaliert werden kann oder Anwendungsfehler verursacht werden.



- Es wurde ein Problem behoben, bei dem Skalierungsanforderungen für einen großen, stark ausgelasteten Cluster fehlschlagen, wenn Amazon-EMR-On-Cluster-Daemons Aktivitäten zur Integritätsprüfung durchführten, z. B. das Erfassen des YARN-Knotenstatus und des HDFS-Knotenstatus. Das lag daran, dass Cluster-Daemons die Integritätsstatusdaten eines Knotens nicht an interne Amazon-EMR-Komponenten weitergeben konnten.
- Verbesserte EMR-On-Cluster-Daemons zur korrekten Nachverfolgung der Knotenstatus bei der Wiederverwendung von IP-Adressen, um die Zuverlässigkeit bei Skalierungsvorgängen zu verbessern.
- [SPARK-29683](#). Es wurde ein Problem behoben, bei dem während der Cluster-Skalierung Auftragsfehler auftraten, da Spark davon ausging, dass alle verfügbaren Knoten auf der Verweigern-Liste standen.
- [YARN-9011](#). Es wurde ein Problem behoben, bei dem Auftragsfehler aufgrund eines Fehlers bei der Außerbetriebnahme von YARN auftraten, wenn der Cluster versuchte, hoch- oder herunterzuskalieren.
- Das Problem mit Schritt- oder Auftragsfehlern bei der Cluster-Skalierung wurde behoben, indem sichergestellt wurde, dass die Knotenstatus zwischen den Amazon-EMR-On-Cluster-Daemons und YARN/HDFS immer konsistent sind.
- Es wurde ein Problem behoben, bei dem Clustervorgänge wie Herunterskalierung und Schrittübermittlung für Amazon-EMR-Cluster, die mit Kerberos-Authentifizierung aktiviert waren, fehlschlagen. Dies lag daran, dass der Amazon-EMR-On-Cluster-Daemon das Kerberos-Ticket nicht erneuert hat, das für die sichere Kommunikation mit HDFS/YARN erforderlich ist, das auf dem Primärknoten ausgeführt wird.
- Neuere Amazon-EMR-Versionen beheben das Problem mit einem niedrigeren Limit für die „maximale Anzahl geöffneter Dateien“ auf älteren AL2-Versionen in Amazon EMR. Die Amazon-EMR-Versionen 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 und höher enthalten jetzt einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“.
- Aktualisierte Komponentenversionen.
- Eine Liste der Komponentenversionen finden Sie in diesem Handbuch unter [Über Amazon-EMR-Versionen](#).

## Neue Features

- Ab Amazon EMR 5.32.0 und 6.5.0 ist die dynamische Executor-Größenanpassung für Apache Spark standardmäßig aktiviert. Sie können dieses Feature ein- oder ausschalten, indem Sie den Konfigurationsparameter `spark.yarn.heterogeneousExecutors.enabled` verwenden.

- Supportstatus für Instance Metadata Service (IMDS) V2: Amazon-EMR-Komponenten von 5.23.1, 5.27.1 und 5.32 oder höher verwenden IMDSv2 für alle IMDS-Aufrufe. Für IMDS-Aufrufe in Ihrem Anwendungscode können Sie sowohl IMDSv1 als auch IMDSv2 verwenden oder das IMDS so konfigurieren, dass es aus Sicherheitsgründen nur IMDSv2 verwendet. Bei anderen 5.x-EMR-Versionen führt die Deaktivierung von IMDSv1 zu einem Cluster-Startup-Fehler.
- Mit Amazon EMR 5.32.0 können Sie einen Cluster starten, der nativ in Apache Ranger integriert ist. Apache Ranger ist ein Open-Source-Framework zur Aktivierung, Überwachung und Verwaltung einer umfassenden Datensicherheit auf der gesamten Hadoop-Plattform. Weitere Informationen finden Sie unter [Apache Ranger](#). Dank der nativen Integration können Sie Ihren eigenen Apache Ranger verwenden, um eine detaillierte Datenzugriffskontrolle auf Amazon EMR durchzusetzen. Weitere Informationen finden Sie unter [Amazon EMR mit Apache Ranger](#) im Verwaltungshandbuch für Amazon EMR.
- Amazon-EMR-Version 5.32.0 unterstützt Amazon EMR in EKS. Weitere Informationen zu den ersten Schritten mit EMR auf EKS finden Sie unter [Was ist Amazon EMR in EKS](#).
- Amazon-EMR-Version 5.32.0 unterstützt Amazon EMR Studio (Vorversion). Weitere Informationen zu den ersten Schritten mit EMR Studio finden Sie unter [Amazon EMR Studio \(Vorversion\)](#).
- Verwaltete Richtlinien mit Geltungsbereich: Um den AWS bewährten Methoden zu entsprechen, hat Amazon EMR verwaltete Standardrichtlinien mit EMR-Geltungsbereich der Version v2 eingeführt, die als Ersatz für Richtlinien dienen, die nicht mehr unterstützt werden. Weitere Informationen finden Sie unter [Verwaltete Richtlinien von Amazon EMR](#).

## Bekannte Probleme

- Für private Subnetz-Cluster von Amazon EMR 6.3.0 und 6.2.0 können Sie nicht auf die Ganglia-Webbenutzeroberfläche zugreifen. Sie erhalten die Fehlermeldung „Zugriff verweigert (403)“. Andere Web-Benutzeroberflächen wie Spark, Hue, Zeppelin, Livy und Tez funktionieren JupyterHub normal. Der Zugriff auf die Ganglia-Web-Benutzeroberfläche auf öffentlichen Subnetzclustern funktioniert ebenfalls normal. Um dieses Problem zu beheben, starten Sie den httpd-Service auf dem Primärknoten mit `sudo systemctl restart httpd` neu. Dieses Problem wurde in Amazon EMR 6.4.0 behoben.
- Niedrigeres Limit für die „Max. Anzahl geöffneter Dateien“ auf älteren AL2-Versionen [in neueren Versionen behoben]. Amazon-EMR-Versionen: emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0 und emr-6.2.0 basieren auf älteren Versionen von Amazon Linux 2 (AL2), die eine niedrigere Ulimit-Einstellung für „Max. Anzahl geöffneter Dateien“ haben, wenn Amazon-EMR-Cluster mit dem Standard-AMI erstellt werden. Die Amazon-EMR-Versionen 5.30.1, 5.30.2, 5.31.1,

5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 und höher enthalten einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“. Versionen mit einem niedrigeren Limit für geöffnete Dateien verursachen beim Senden des Spark-Jobs den Fehler „Zu viele offene Dateien“. In den betroffenen Versionen hat das Amazon-EMR-Standard-AMI eine Ulimit-StandardEinstellung von 4 096 für „Max. Anzahl geöffneter Dateien“, was unter dem Dateilimit von 65 536 im neuesten Amazon-Linux-2-AMI liegt. Die niedrigere Ulimit-Einstellung für „Max. Anzahl geöffneter Dateien“ führt dazu, dass der Spark-Job fehlschlägt, wenn der Spark-Treiber und der Executor versuchen, mehr als 4 096 Dateien zu öffnen. Um das Problem zu beheben, verfügt Amazon EMR über ein Bootstrap Action (BA)-Skript, das die Ulimit-Einstellung bei der Cluster-Erstellung anpasst.

Wenn Sie eine ältere Amazon-EMR-Version verwenden, für die dieses Problem nicht dauerhaft behoben ist, können Sie mit der folgenden Problemumgehung das Instance-Controller-Ulimit explizit auf maximal 65 536 Dateien festlegen.

Explizit ein ulimit über die Befehlszeile setzen

1. Bearbeiten Sie `/etc/systemd/system/instance-controller.service`, um die folgenden Parameter zum Abschnitt Service hinzuzufügen.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Starten Sie neu InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Mithilfe der Bootstrap-Aktion (BA) ein Ulimit festlegen

Sie können auch ein Bootstrap-Aktionsskript (BA) verwenden, um das Ulimit für den Instance-Controller bei der Clustererstellung auf 65 536 Dateien zu konfigurieren.

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
```

```
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

### Important

EMR-Cluster, auf denen Amazon-Linux- oder Amazon-Linux-2-AMIs (Amazon Machine Images) ausgeführt werden, verwenden das Standardverhalten von Amazon Linux und laden wichtige und kritische Kernel-Updates, die einen Neustart erfordern, nicht automatisch herunter und installieren sie. Dies ist dasselbe Verhalten wie bei anderen Amazon-EC2-Instances, die das standardmäßige Amazon-Linux-AMI ausführen. Wenn neue Amazon-Linux-Softwareupdates, die einen Neustart erfordern (wie Kernel-, NVIDIA- und CUDA-Updates), nach der Veröffentlichung einer Amazon-EMR-Version verfügbar werden, laden EMR-Cluster-Instances, die das Standard-AMI ausführen, diese Updates nicht automatisch herunter und installieren sie. Um Kernel-Updates zu erhalten, können Sie [Ihr Amazon-EMR-AMI](#) so anpassen, dass es [das neueste Amazon-Linux-AMI verwendet](#).

- Die Konsolenunterstützung zur Erstellung einer Sicherheitskonfiguration, die die AWS Ranger-Integrationsoption spezifiziert, wird derzeit in der GovCloud Region nicht unterstützt. Die Sicherheitskonfiguration kann mit der CLI durchgeführt werden. Weitere Informationen finden [Sie unter Erstellen der EMR-Sicherheitskonfiguration](#) im Verwaltungshandbuch für Amazon EMR.
- Wenn AtRestEncryption oder die HDFS-Verschlüsselung auf einem Cluster aktiviert ist, der Amazon EMR 5.31.0 oder 5.32.0 verwendet, führen Hive-Abfragen zu der folgenden Laufzeitausnahme.

```
TaskAttempt 3 failed, info=[Error: Error while running task ( failure ) :
attempt_1604112648850_0001_1_01_000000_3:java.lang.RuntimeException:
java.lang.RuntimeException: Hive Runtime Error while closing
operators: java.io.IOException: java.util.ServiceConfigurationError:
org.apache.hadoop.security.token.TokenIdentifier: Provider
org.apache.hadoop.hbase.security.token.AuthenticationTokenIdentifier not found
```

- Wenn Sie Spark mit der Formatierung des Hive-Partitionsstandorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den Amazon-EMR-Versionen 5.30.0 bis 5.36.0 und 6.2.0 bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:
  - Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.
  - Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist `s3://bucket/table/p=a` ein Präfix von `s3://bucket/table/p=a b`.
  - Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen UTF-8-Wert, der kleiner als das `/`-Zeichen (U+002F) ist. Beispielsweise fällt das Leerzeichen (U+0020), das in `s3://bucket/table/p=a b` zwischen `a` und `b` vorkommt, in diese Kategorie. Beachten Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: `!"#$%&'()*+,-.` Weitere Informationen finden Sie unter [UTF-8-Kodierungstabelle und Unicode-Zeichen](#).

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.

## 5.32.0 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-`amzn-EmrVersion`. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1.4.1	Amazon SageMaker Spark-SDK
emr-ddb	4.16.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.13.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.5.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-notebook-env	1.1.0	Conda Env für EMR-Notebooks, das Jupyter Enterprise Gateway enthält
emr-s3-dist-cp	2.17.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	1.6.0	EMR S3Select-Konnektor
emrfs	2,45,0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.11.2	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
flink-jobmanager-config	1.11.2	Verwaltung von Ressourcen auf EMR-Knoten für Apache JobManager Flink.

Komponente	Version	Beschreibung
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.10.1-amzn-0	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.10.1-amzn-0	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.10.1-amzn-0	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.10.1-amzn-0	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-hdfs-journalnode	2.10.1-amzn-0	HDFS-Service zum Verwalten des Hadoop-Dateisystem journals auf HA-Clustern.

Komponente	Version	Beschreibung
hadoop-https-server	2.10.1-amzn-0	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.10.1-amzn-0	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.10.1-amzn-0	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.10.1-amzn-0	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resource-manager	2.10.1-amzn-0	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.10.1-amzn-0	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.4.13	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.4.13	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.



Komponente	Version	Beschreibung
hbase-client	1.4.13	HBase-Befehlszeilen-Client.
hbase-rest-server	1.4.13	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.4.13	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.7-amzn-3	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.7-amzn-3	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.3.7-amzn-3	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.3.7-amzn-3	Hive-Befehlszeilen-Client.
hive-hbase	2.3.7-amzn-3	Hive-hbase client.
hive-metastore-server	2.3.7-amzn-3	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.3.7-amzn-3	Service zur Annahme von Hive-Abfragen als Webanfragen.

Komponente	Version	Beschreibung
hudi	0.6.0-amzn-0	Inkrementelles Verarbeitungs-Framework zur Stromversorgung der Datenpipeline bei geringer Latenz und hoher Effizienz.
hudi-spark	0.6.0-amzn-0	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.
hudi-presto	0.6.0-amzn-0	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hue-server	4.8.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	1.1.0	Multi-User-Server für Jupyter-Notebooks
livy-server	0.7.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP- und Reverse-Proxy-Server.
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.7.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.68	MySQL-Datenbankserver.
nvidia-cuda	10,1243	Nvidia-Treiber und Cuda-Tool kit

Komponente	Version	Beschreibung
oozie-client	5.2.0	Oozie-Befehlszeilen-Client.
oozie-server	5.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	4.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.3-HBase-1.4	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.14.3-HBase-1.4	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.240.1-amzn-0	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.240.1-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.240.1-amzn-0	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.

Komponente	Version	Beschreibung
r	3.4.3	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	1.2.0	Apache Ranger Key Management System
spark-client	2.4.7-amzn-0	Spark-Befehlszeilen-Clients.
spark-history-server	2.4.7-amzn-0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.4.7-amzn-0	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.4.7-amzn-0	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	2.3.1	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.2	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.

Komponente	Version	Beschreibung
zeppelin-server	0.8.2	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.14	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.14	ZooKeeper Befehlszeilen-Cli-ent.

## 5.32.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Aktionen zur Neukonfiguration treten auf, wenn Sie eine Konfiguration für Instance-Gruppen in einem laufenden Cluster angeben. Amazon EMR initiiert nur Rekonfigurationsaktionen für die Klassifizierungen, die Sie ändern. Weitere Informationen finden Sie unter [Eine Instance-Gruppe in einem laufenden Cluster neu konfigurieren](#).

### emr-5.32.0-Klassifizierungen

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.	Restarts the Resource Manager service.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
container-executor	Ändern Sie die Werte in der Datei „container-executor.cfg“ Datei von Hadoop YARN.	Not available.
container-log4j	Ändert die Werte in der container-log4j.properties-Datei in Hadoop YARN.	Not available.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Ändern Sie die Docker-bezogenen Einstellungen.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
emrfs-site	Ändert die EMRFS-Einstellungen.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegistrator, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.	Restarts Flink history server.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.	Restarts Flink history server.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4j-yarn-session.properties.	Not available.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.	Restarts Flink history server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.	Not available.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.	Not available.



Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Ändert die Werte in der HBase-Umgebung.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.	Not available.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.	Should not be reconfigured.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.	Restarts Hive WebHCat Server.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.	Restarts Hive WebHCat Server.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.	Restarts Hive WebHCat Server.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.	Not available.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.	Not available.
hive-env	Ändert die Werte in der Hive-Umgebung.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.	Restarts HiveServer2 and HiveMetastore.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log	Not available.
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.	Not available.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.	Not available.
hue-ini	Ändert die Werte in der INI-Datei in Hue.	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.	Restarts Hadoop Httpfs service.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.	Not available.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.	Restarts Hadoop-KMS service.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.	Not available.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Ändern der Werte in der Hudi-Umgebung.	Not available.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.	Not available.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.	Not available.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.	Not available.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.	Not available.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.	Restarts Livy Server.
livy-env	Ändert die Werte in der Livy-Umgebung.	Restarts Livy Server.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.	Restarts Livy Server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
mapred-env	Ändern Sie die Werte in der MapReduce Anwendung sumgebung.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Ändert die Werte in der Oozie-Umgebung.	Restarts Oozie.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.	Restarts Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.	Not available.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.	Not available.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.	Not available.
pig-env	Ändert die Werte in der Pig-Umgebung.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.	Restarts Oozie.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.	Not available.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.	Restarts Presto-Server.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.	Restarts Presto-Server.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.	Not available.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.	Restarts Presto-Server.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.	Not available.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.	Not available.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.	Not available.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.	Restarts Presto-Server.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.	Not available.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.	Not available.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.	Not available.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.	Not available.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.	Not available.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.	Not available.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.	Not available.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.	Not available.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.	Not available.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.	Not available.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.	Not available.
ranger-kms-dbks-site	Ändert die Werte in der dbks-site.xml-Datei von Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger KMS.	Restarts Ranger KMS Server.
ranger-kms-env	Ändert die Werte in der Ranger KMS-Umgebung.	Restarts Ranger KMS Server.
ranger-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei von Ranger KMS.	Not available.
ranger-kms-db-ca	Ändert die Werte für die CA-Datei auf S3 für die MySQL SSL-Verbindung mit Ranger KMS.	Not available.
recordserver-env	Werte in der RecordServer EMR-Umgebung ändern.	Restarts EMR record server.
recordserver-conf	Ändern Sie die Werte in der Datei RecordServer server.properties von EMR.	Restarts EMR record server.



Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
recordserver-log4j	Ändern Sie die Werte in der Datei RecordServer log4j.properties von EMR.	Restarts EMR record server.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Ändert die Werte in der Spark-Umgebung.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.	Not available.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.	Not available.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml von Sqoop OraOop.	Not available.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.	Not available.

Klassifizierungen	Beschreibung	Aktionen zur Neukonfiguration
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.	Restarts Oozie.
yarn-env	Ändert die Werte in der YARN-Umgebung.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.	Restarts Zeppelin.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.	Restarts Zookeeper server.

## Amazon-EMR-Version 5.31.1

### 5.31.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

#### Informationen zur Anwendungsversion

	emr-5.31.1	emr-5.31.0	emr-5.30.2	emr-5.30.1
AWS SDK for Java	1.11.852	1,11.852	1,11.759	1,11.759
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.11.0	1.11.0	1.10.0	1.10.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.13	1.4.13	1.4.13	1.4.13
HCatalog	2.3.7	2.3.7	2.3.6	2.3.6
Hadoop	2.10.0	2.10.0	2.8.5	2.8.5
Hive	2.3.7	2.3.7	2.3.6	2.3.6

	emr-5.31.1	emr-5.31.0	emr-5.30.2	emr-5.30.1
Hudi	0.6.0-amzn-0	0.6.0-amzn-0	0.5.2-incubating	0.5.2-incubating
Hue	4.7.1	4.7.1	4.6.0	4.6.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.1.0	1.1.0	1.1.0	1.1.0
Livy	0.7.0	0.7.0	0.7.0	0.7.0
MXNet	1.6.0	1.6.0	1.5.1	1.5.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.0	5.2.0	5.2.0	5.2.0
Phoenix	4.14.3	4.14.3	4.14.3	4.14.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0,238,3	0,238,3	0.232	0.232
Spark	2.4.6	2.4.6	2.4.5	2.4.5
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.1.0	2.1.0	1.14.0	1.14.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.2	0.8.2	0.8.2	0.8.2
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

## 5.31.1 Versionshinweise

Dies ist eine Version zur Behebung von Problemen mit Amazon-EMR-Skalierung, wenn ein Cluster nicht erfolgreich hoch-/herunterskaliert werden kann oder Anwendungsfehler verursacht werden.

### Änderungen, Verbesserungen und behobene Probleme

- Es wurde ein Problem behoben, bei dem Skalierungsanforderungen für einen großen, stark ausgelasteten Cluster fehlschlagen, wenn Amazon-EMR-On-Cluster-Daemons Aktivitäten zur Integritätsprüfung durchführten, z. B. das Erfassen des YARN-Knotenstatus und des HDFS-Knotenstatus. Das lag daran, dass Cluster-Daemons die Integritätsstatusdaten eines Knotens nicht an interne Amazon-EMR-Komponenten weitergeben konnten.
- Verbesserte EMR-On-Cluster-Daemons zur korrekten Nachverfolgung der Knotenstatus bei der Wiederverwendung von IP-Adressen, um die Zuverlässigkeit bei Skalierungsvorgängen zu verbessern.
- [SPARK-29683](#). Es wurde ein Problem behoben, bei dem während der Cluster-Skalierung Auftragsfehler auftraten, da Spark davon ausging, dass alle verfügbaren Knoten auf der Verweigern-Liste standen.
- [YARN-9011](#). Es wurde ein Problem behoben, bei dem Auftragsfehler aufgrund eines Fehlers bei der Außerbetriebnahme von YARN auftraten, wenn der Cluster versuchte, hoch- oder herunterzuskalieren.
- Das Problem mit Schritt- oder Auftragsfehlern bei der Cluster-Skalierung wurde behoben, indem sichergestellt wurde, dass die Knotenstatus zwischen den Amazon-EMR-On-Cluster-Daemons und YARN/HDFS immer konsistent sind.
- Es wurde ein Problem behoben, bei dem Clustervorgänge wie Herunterskalierung und Schrittübermittlung für Amazon-EMR-Cluster, die mit Kerberos-Authentifizierung aktiviert waren, fehlschlagen. Dies lag daran, dass der Amazon-EMR-On-Cluster-Daemon das Kerberos-Ticket nicht erneuert hat, das für die sichere Kommunikation mit HDFS/YARN erforderlich ist, das auf dem Primärknoten ausgeführt wird.
- Neuere Amazon-EMR-Versionen beheben das Problem mit einem niedrigeren Limit für die „maximale Anzahl geöffneter Dateien“ auf älteren AL2-Versionen in Amazon EMR. Die Amazon-EMR-Versionen 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 und höher enthalten jetzt einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“.
- HTTPS ist jetzt standardmäßig für Amazon-Linux-Repositorys aktiviert. Wenn Sie eine Amazon-S3-VPCE-Richtlinie verwenden, um den Zugriff auf bestimmte Buckets zu beschränken, müssen Sie den neuen Amazon-Linux-Bucket-ARN `arn:aws:s3:::amazonlinux-2-repos-$region/*`

zu Ihrer Richtlinie hinzufügen (ersetzen Sie `$region` durch die Region, in der sich der Endpunkt befindet). Weitere Informationen finden Sie unter diesem Thema in den AWS Diskussionsforen. [Ankündigung: Amazon Linux 2 unterstützt jetzt die Möglichkeit, HTTPS zu verwenden, wenn eine Verbindung zu Paket-Repositorys hergestellt wird.](#)

## Bekannte Probleme

- Wenn Sie Spark mit der Formatierung des Hive-Partitionsstandorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den Amazon-EMR-Versionen 5.30.0 bis 5.36.0 und 6.2.0 bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:
  - Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.
  - Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist `s3://bucket/table/p=a` ein Präfix von `s3://bucket/table/p=a b`.
  - Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen UTF-8-Wert, der kleiner als das `/`-Zeichen (U+002F) ist. Beispielsweise fällt das Leerzeichen (U+0020), das in `s3://bucket/table/p=a b` zwischen `a` und `b` vorkommt, in diese Kategorie. Beachten Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: `!"#$%&'()*+,-.` Weitere Informationen finden Sie unter [UTF-8-Kodierungstabelle und Unicode-Zeichen](#).

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.

### 5.31.1 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn zum Beispiel eine Open-Source-

Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.4.0	Amazon SageMaker Spark-SDK
<code>emr-ddb</code>	4.15.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.13.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.5.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-s3-dist-cp</code>	2.15.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
<code>emr-s3-select</code>	1.6.0	EMR S3Select-Konnektor
<code>emrfs</code>	2,43,0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>flink-client</code>	1.11.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
<code>flink-jobmanager-config</code>	1.11.0	Verwaltung von Ressourcen auf EMR-Knoten für Apache JobManager Flink.

Komponente	Version	Beschreibung
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.10.0-amzn-0.1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.10.0-amzn-0.1	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.10.0-amzn-0.1	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.10.0-amzn-0.1	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-hdfs-journalnode	2.10.0-amzn-0.1	HDFS-Service zum Verwalten des Hadoop-Dateisystem journals auf HA-Clustern.



Komponente	Version	Beschreibung
hadoop-httfs-server	2.10.0-amzn-0.1	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.10.0-amzn-0.1	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.10.0-amzn-0.1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.10.0-amzn-0.1	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.10.0-amzn-0.1	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.10.0-amzn-0.1	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.4.13	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.4.13	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.

Komponente	Version	Beschreibung
hbase-client	1.4.13	HBase-Befehlszeilen-Client.
hbase-rest-server	1.4.13	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.4.13	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.7-amzn-1	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.7-amzn-1	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.3.7-amzn-1	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.3.7-amzn-1	Hive-Befehlszeilen-Client.
hive-hbase	2.3.7-amzn-1	Hive-hbase client.
hive-metastore-server	2.3.7-amzn-1	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.3.7-amzn-1	Service zur Annahme von Hive-Abfragen als Webanfragen.

Komponente	Version	Beschreibung
hudi	0.6.0-amzn-0	Inkrementelles Verarbeitungs-Framework zur Stromversorgung der Datenpipeline bei geringer Latenz und hoher Effizienz.
hudi-spark	0.6.0-amzn-0	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.
hudi-presto	0.6.0-amzn-0	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hue-server	4.7.1	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	1.1.0	Multi-User-Server für Jupyter-Notebooks
livy-server	0.7.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP- und Reverse-Proxy-Server.
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.6.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.64+	MySQL-Datenbankserver.
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Toolkit

Komponente	Version	Beschreibung
oozie-client	5.2.0	Oozie-Befehlszeilen-Client.
oozie-server	5.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	4.3.0	Open Source Computer Vision Library.
phoenix-library	4.14.3-HBase-1.4	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.14.3-HBase-1.4	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.238.3-amzn-0	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.238.3-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.238.3-amzn-0	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.

Komponente	Version	Beschreibung
r	3.4.3	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	1.2.0	Apache Ranger Key Management System
spark-client	2.4.6-amzn-0.1	Spark-Befehlszeilen-Clients.
spark-history-server	2.4.6-amzn-0.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.4.6-amzn-0.1	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.4.6-amzn-0.1	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	2.1.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.2	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.

Komponente	Version	Beschreibung
zeppelin-server	0.8.2	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.14	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.14	ZooKeeper Befehlszeilen-Cli-ent.

### 5.31.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

#### emr-5.31.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-log4j	Ändert die Werte in der <code>container-log4j.properties</code> -Datei in Hadoop YARN.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.

Klassifizierungen	Beschreibung
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4 j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.

Klassifizierungen	Beschreibung
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.



Klassifizierungen	Beschreibung
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
hudi-env	Ändern der Werte in der Hudi-Umgebung.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.

Klassifizierungen	Beschreibung
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.

Klassifizierungen	Beschreibung
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.
ranger-kms-dbks-site	Ändert die Werte in der dbks-site.xml-Datei von Ranger KMS.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger KMS.
ranger-kms-env	Ändert die Werte in der Ranger KMS-Umgebung.
ranger-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei von Ranger KMS.
ranger-kms-db-ca	Ändert die Werte für die CA-Datei auf S3 für die MySQL SSL-Verbindung mit Ranger KMS.

Klassifizierungen	Beschreibung
recordserver-env	Werte in der RecordServer EMR-Umgebung ändern.
recordserver-conf	Ändern Sie die Werte in der Datei RecordServer server.properties von EMR.
recordserver-log4j	Ändern Sie die Werte in der Datei RecordServer log4j.properties von EMR.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml von Sqoop OraOop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.

Klassifizierungen	Beschreibung
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.31.0

### 5.31.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

## Informationen zur Anwendungsversion

	emr-5.31.0	emr-5.30.2	emr-5.30.1	emr-5.30.0
AWS SDK for Java	1.11.852	1,11.759	1,11.759	1,11.759
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.11.0	1.10.0	1.10.0	1.10.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.13	1.4.13	1.4.13	1.4.13
HCatalog	2.3.7	2.3.6	2.3.6	2.3.6
Hadoop	2.10.0	2.8.5	2.8.5	2.8.5
Hive	2.3.7	2.3.6	2.3.6	2.3.6
Hudi	0.6.0-amzn-0	0.5.2-incubating	0.5.2-incubating	0.5.2-incubating
Hue	4.7.1	4.6.0	4.6.0	4.6.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.1.0	1.1.0	1.1.0	1.1.0
Livy	0.7.0	0.7.0	0.7.0	0.7.0
MXNet	1.6.0	1.5.1	1.5.1	1.5.1

	emr-5.31.0	emr-5.30.2	emr-5.30.1	emr-5.30.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.0	5.2.0	5.2.0	5.2.0
Phoenix	4.14.3	4.14.3	4.14.3	4.14.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0,238,3	0.232	0.232	0.232
Spark	2.4.6	2.4.5	2.4.5	2.4.5
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.1.0	1.14.0	1.14.0	1.14.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.2	0.8.2	0.8.2	0.8.2
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

## 5.31.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.31.0. Änderungen beziehen sich auf Version 5.30.1.

Erste Version: 09. Oktober 2020

Letzte Aktualisierung: 15. Oktober 2020

### Upgrades

- Der Amazon-Glue-Konnektor wurde auf Version 1.13.0 aktualisiert
- Amazon SageMaker Spark SDK wurde auf Version 1.4.0 aktualisiert
- Der Amazon-Kinesis-Konnektor wurde auf Version 3.5.9 aktualisiert



- Auf Version AWS SDK for Java 1.11.852 aktualisiert
- BigTop-Tomcat wurde auf Version 8.5.56 aktualisiert
- EMR FS wurde auf Version 2.43.0 aktualisiert
- MetricsAndEventsApiGateway EMR-Client wurde auf Version 1.4.0 aktualisiert
- EMR S3 Dist CP wurde auf Version 2.15.0 aktualisiert
- EMR S3 Select wurde auf Version 1.6.0 aktualisiert
- Flink wurde auf Version 1.11.0 aktualisiert
- Hadoop wurde auf Version 2.10.0 aktualisiert
- Hive wurde auf Version 2.3.7 aktualisiert
- Hudi wurde auf Version 0.6.0 aktualisiert
- Hue wurde auf Version 4.7.1 aktualisiert
- Auf Version JupyterHub 1.1.0 aktualisiert
- Mxnet wurde auf Version 1.6.0 aktualisiert
- OpenCV wurde auf Version 4.3.0 aktualisiert
- Presto wurde auf Version 0.238.3 aktualisiert
- Auf Version TensorFlow 2.1.0 aktualisiert

## Änderungen, Verbesserungen und behobene Probleme

- Dies ist eine Version zur Behebung von Problemen mit Amazon-EMR-Skalierung, wenn ein Cluster nicht erfolgreich hoch-/herunterskaliert werden kann oder Anwendungsfehler verursacht werden.
- Es wurde ein Problem behoben, bei dem Skalierungsanforderungen für einen großen, stark ausgelasteten Cluster fehlschlagen, wenn Amazon-EMR-On-Cluster-Daemons Aktivitäten zur Integritätsprüfung durchführten, z. B. das Erfassen des YARN-Knotenstatus und des HDFS-Knotenstatus. Das lag daran, dass Cluster-Daemons die Integritätsstatusdaten eines Knotens nicht an interne Amazon-EMR-Komponenten weitergeben konnten.
- Verbesserte EMR-On-Cluster-Daemons zur korrekten Nachverfolgung der Knotenstatus bei der Wiederverwendung von IP-Adressen, um die Zuverlässigkeit bei Skalierungsvorgängen zu verbessern.
- [SPARK-29683](#). Es wurde ein Problem behoben, bei dem während der Cluster-Skalierung Auftragsfehler auftraten, da Spark davon ausging, dass alle verfügbaren Knoten auf der Verweigern-Liste standen.

- [YARN-9011](#). Es wurde ein Problem behoben, bei dem Auftragsfehler aufgrund eines Fehlers bei der Außerbetriebnahme von YARN auftraten, wenn der Cluster versuchte, hoch- oder herunterzuskalieren.
- Das Problem mit Schritt- oder Auftragsfehlern bei der Cluster-Skalierung wurde behoben, indem sichergestellt wurde, dass die Knotenstatus zwischen den Amazon-EMR-On-Cluster-Daemons und YARN/HDFS immer konsistent sind.
- Es wurde ein Problem behoben, bei dem Clustervorgänge wie Herunterskalierung und Schrittübermittlung für Amazon-EMR-Cluster, die mit Kerberos-Authentifizierung aktiviert waren, fehlschlagen. Dies lag daran, dass der Amazon-EMR-On-Cluster-Daemon das Kerberos-Ticket nicht erneuert hat, das für die sichere Kommunikation mit HDFS/YARN erforderlich ist, das auf dem Primärknoten ausgeführt wird.
- Neuere Amazon-EMR-Versionen beheben das Problem mit einem niedrigeren Limit für die „maximale Anzahl geöffneter Dateien“ auf älteren AL2-Versionen in Amazon EMR. Die Amazon-EMR-Versionen 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 und höher enthalten jetzt einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“.
- [Hive Spaltenstatistiken](#) werden für Amazon-EMR-5.31.0 und höher unterstützt.
- Aktualisierte Komponentenversionen.
- Support für EMRFS S3EC V2 in Amazon EMR 5.31.0. In den S3-Java-SDK-Versionen 1.11.837 und höher wurde der Verschlüsselungsclient Version 2 (S3EC V2) mit verschiedenen Sicherheitsverbesserungen eingeführt. Weitere Informationen finden Sie hier:
  - S3 Blogbeitrag: [Updates für den Amazon S3 Encryption Client](#).
  - AWS SDK for Java Entwicklerhandbuch: [Migrieren Sie Verschlüsselungs- und Entschlüsselungsclients auf V2](#).
  - EMR-Verwaltungshandbuch: [Amazon-S3-Clientseitige Verschlüsselung](#).

Der Encryption Client V1 ist aus Gründen der Abwärtskompatibilität weiterhin im SDK verfügbar.

## Neue Features

- Niedrigeres Limit für die „Max. Anzahl geöffneter Dateien“ auf älteren AL2-Versionen [in neueren Versionen behoben]. Amazon-EMR-Versionen: emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0 und emr-6.2.0 basieren auf älteren Versionen von Amazon Linux 2 (AL2), die eine niedrigere Ulimit-Einstellung für „Max. Anzahl geöffneter Dateien“ haben, wenn Amazon-EMR-Cluster mit dem Standard-AMI erstellt werden. Die Amazon-EMR-Versionen 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 und höher enthalten einen dauerhaften Fix mit einer

höheren Einstellung für „Max. Anzahl geöffneter Dateien“. Versionen mit einem niedrigeren Limit für geöffnete Dateien verursachen beim Senden des Spark-Jobs den Fehler „Zu viele offene Dateien“. In den betroffenen Versionen hat das Amazon-EMR-Standard-AMI eine Ulimit-StandardEinstellung von 4 096 für „Max. Anzahl geöffneter Dateien“, was unter dem Dateilimit von 65 536 im neuesten Amazon-Linux-2-AMI liegt. Die niedrigere Ulimit-Einstellung für „Max. Anzahl geöffneter Dateien“ führt dazu, dass der Spark-Job fehlschlägt, wenn der Spark-Treiber und der Executor versuchen, mehr als 4 096 Dateien zu öffnen. Um das Problem zu beheben, verfügt Amazon EMR über ein Bootstrap Action (BA)-Skript, das die Ulimit-Einstellung bei der Cluster-Erstellung anpasst.

Wenn Sie eine ältere Amazon-EMR-Version verwenden, für die dieses Problem nicht dauerhaft behoben ist, können Sie mit der folgenden Problemumgehung das Instance-Controller-Ulimit explizit auf maximal 65 536 Dateien festlegen.

Explizit ein ulimit über die Befehlszeile setzen

1. Bearbeiten Sie `/etc/systemd/system/instance-controller.service`, um die folgenden Parameter zum Abschnitt Service hinzuzufügen.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Starten Sie neu InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Mithilfe der Bootstrap-Aktion (BA) ein Ulimit festlegen

Sie können auch ein Bootstrap-Aktionsskript (BA) verwenden, um das Ulimit für den Instance-Controller bei der Clustererstellung auf 65 536 Dateien zu konfigurieren.

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
```

```
sudo mkdir -p /etc/systemd/system/$proc.service.d/  
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF  
[Service]  
LimitNOFILE=65536  
LimitNPROC=65536  
EOF  
pid=$(pgrep -f aws157.$proc.Main)  
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535  
done  
sudo systemctl daemon-reload
```

- Mit Amazon EMR 5.31.0 können Sie einen Cluster starten, der in Lake Formation integriert ist. Diese Integration ermöglicht eine detaillierte Datenfilterung auf Spaltenebene für Datenbanken und Tabellen im Glue-Datenkatalog. AWS Es ermöglicht Lake Formation über ein Unternehmens-Identitätssystem eine verbundene einmalige Anmeldung bei EMR Notebooks oder Apache Zeppelin. Weitere Informationen über [Integrieren von Amazon EMR mit AWS Lake Formation](#) finden Sie im Verwaltungshandbuch für Amazon EMR.

Amazon EMR with Lake Formation ist derzeit in 16 AWS Regionen verfügbar: USA Ost (Ohio und Nord-Virginia), USA West (Nordkalifornien und Oregon), Asien-Pazifik (Mumbai, Seoul, Singapur, Sydney und Tokio), Kanada (Zentral), Europa (Frankfurt, Irland, London, Paris und Stockholm), Südamerika (São Paulo).

## Bekannte Probleme

- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in Amazon-EMR-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie dem Herunterskalieren oder der schrittweisen Übermittlung auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

### Workaround:

- SSH als hadoop-Benutzer für den führenden Primärknoten des EMR-Clusters mit mehreren Primärknoten.

- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den hadoop-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

#### Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

- Wenn AtRestEncryption oder die HDFS-Verschlüsselung auf einem Cluster aktiviert ist, der Amazon EMR 5.31.0 oder 5.32.0 verwendet, führen Hive-Abfragen zu der folgenden Laufzeitausnahme.

```
TaskAttempt 3 failed, info=[Error: Error while running task ( failure ) :
attempt_1604112648850_0001_1_01_000000_3:java.lang.RuntimeException:
java.lang.RuntimeException: Hive Runtime Error while closing
operators: java.io.IOException: java.util.ServiceConfigurationError:
org.apache.hadoop.security.token.TokenIdentifier: Provider
org.apache.hadoop.hbase.security.token.AuthenticationTokenIdentifier not found
```

- Wenn Sie Spark mit der Formatierung des Hive-Partitionsstandorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den Amazon-EMR-Versionen 5.30.0 bis 5.36.0 und 6.2.0 bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:
  - Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.
  - Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist `s3://bucket/table/p=a` ein Präfix von `s3://bucket/table/p=a b`.
  - Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen UTF-8-Wert, der kleiner als das `/`-Zeichen (U+002F) ist. Beispielsweise fällt das Leerzeichen (U+0020), das in `s3://bucket/table/p=a b` zwischen `a` und `b` vorkommt, in diese Kategorie. Beachten

Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: !"#%&'()\*+,-. Weitere Informationen finden Sie unter [UTF-8-Kodierungstabelle und](#) Unicode-Zeichen.

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.

### 5.31.0 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.4.0	Amazon SageMaker Spark-SDK
<code>emr-ddb</code>	4.15.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.13.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.5.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-s3-dist-cp	2.15.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	1.6.0	EMR S3Select-Konnektor
emrfs	2,43,0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.11.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
flink-jobmanager-config	1.11.0	Verwaltung von Ressourcen auf EMR-Knoten für Apache JobManager Flink.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.10.0-amzn-0	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".

Komponente	Version	Beschreibung
hadoop-hdfs-datanode	2.10.0-amzn-0	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.10.0-amzn-0	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.10.0-amzn-0	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-hdfs-journalnode	2.10.0-amzn-0	HDFS-Service zum Verwalten des Hadoop-Dateisystem journals auf HA-Clustern.
hadoop-httfs-server	2.10.0-amzn-0	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.10.0-amzn-0	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.10.0-amzn-0	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.10.0-amzn-0	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.10.0-amzn-0	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.



Komponente	Version	Beschreibung
hadoop-yarn-timeline-server	2.10.0-amzn-0	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.4.13	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.4.13	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.4.13	HBase-Befehlszeilen-Client.
hbase-rest-server	1.4.13	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.4.13	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.7-amzn-1	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.7-amzn-1	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.3.7-amzn-1	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.

Komponente	Version	Beschreibung
hive-client	2.3.7-amzn-1	Hive-Befehlszeilen-Client.
hive-hbase	2.3.7-amzn-1	Hive-hbase client.
hive-metastore-server	2.3.7-amzn-1	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.3.7-amzn-1	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.6.0-amzn-0	Inkrementelles Verarbeitungs-Framework zur Stromversorgung der Datenpipeline bei geringer Latenz und hoher Effizienz.
hudi-spark	0.6.0-amzn-0	Bündel-Bibliothek zum Ausführen von Spark mit Hudi.
hudi-presto	0.6.0-amzn-0	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hue-server	4.7.1	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	1.1.0	Multi-User-Server für Jupyter-Notebooks
livy-server	0.7.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark

Komponente	Version	Beschreibung
nginx	1.12.1	nginx [engine x] ist ein HTTP- und Reverse-Proxy-Server.
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.6.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.64	MySQL-Datenbankserver.
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.2.0	Oozie-Befehlszeilen-Client.
oozie-server	5.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	4.3.0	Open Source Computer Vision Library.
phoenix-library	4.14.3-HBase-1.4	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.14.3-HBase-1.4	Ein schlanker Server für den Zugriff auf JDBC und Protokoll puffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.238.3-amzn-0	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.

Komponente	Version	Beschreibung
presto-worker	0.238.3-amzn-0	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.238.3-amzn-0	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.3	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	1.2.0	Apache Ranger Key Management System
spark-client	2.4.6-amzn-0	Spark-Befehlszeilen-Clients.
spark-history-server	2.4.6-amzn-0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.4.6-amzn-0	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.4.6-amzn-0	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.

Komponente	Version	Beschreibung
tensorflow	2.1.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.2	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.8.2	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.14	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.14	ZooKeeper Befehlszeilen-Cli-ent.

### 5.31.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

## emr-5.31.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der capacity-scheduler.xml-Datei in Hadoop.
container-log4j	Ändert die Werte in der container-log4j.properties-Datei in Hadoop YARN.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4j-yarn-session.properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.

Klassifizierungen	Beschreibung
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.

Klassifizierungen	Beschreibung
hive-parquet-logging	Ändert die Werte in der <code>parquet-logging.properties</code> -Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von <code>hive-exec-log</code> Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von Hive. <code>llap-daemon-log</code>
hive-log4j2	Ändert die Werte in der <code>hive-log4j2.properties</code> -Datei in Hive.
hive-site	Ändert die Werte in der <code>hive-site.xml</code> -Datei in Hive.
hiveserver2-site	Ändert die Werte in der <code>hiveserver2-site.xml</code> -Datei von <code>Server2</code> in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der <code>httpfs-site.xml</code> -Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der <code>kms-acls.xml</code> -Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der <code>kms-log4j.properties</code> -Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der <code>kms-site.xml</code> -Datei in Hadoop.



Klassifizierungen	Beschreibung
hudi-env	Ändern der Werte in der Hudi-Umgebung.
jupyter-notebook-conf	Ändert die Werte in der <code>jupyter_notebook_config.py</code> -Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei <code>jupyterhub_config.py</code> .
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der <code>config.json</code> -Datei in Sparkmagic.
livy-conf	Ändert die Werte in der <code>livy.conf</code> -Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei <code>mapred-site.xml</code> der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der <code>oozie-log4j.properties</code> -Datei in Oozie.
oozie-site	Ändert die Werte in der <code>oozie-site.xml</code> -Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der <code>hadoop-metrics2-hbase.properties</code> -Datei in Phoenix.

Klassifizierungen	Beschreibung
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.

Klassifizierungen	Beschreibung
ranger-kms-dbks-site	Ändert die Werte in der dbks-site.xml-Datei von Ranger KMS.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger KMS.
ranger-kms-env	Ändert die Werte in der Ranger KMS-Umgebung.
ranger-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei von Ranger KMS.
ranger-kms-db-ca	Ändert die Werte für die CA-Datei auf S3 für die MySQL SSL-Verbindung mit Ranger KMS.
recordserver-env	Werte in der RecordServer EMR-Umgebung ändern.
recordserver-conf	Ändern Sie die Werte in der Datei RecordServer server.properties von EMR.
recordserver-log4j	Ändern Sie die Werte in der Datei RecordServer log4j.properties von EMR.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.

Klassifizierungen	Beschreibung
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml von Sqoop OraOop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.30.2

### 5.30.2 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

#### Informationen zur Anwendungsversion

	emr-5.30.2	emr-5.30.1	emr-5.30.0	emr-5.29.0
AWS SDK for Java	1.11.759	1,11.759	1,11.759	1,11,682
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,6
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.10.0	1.10.0	1.10.0	1.9.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.13	1.4.13	1.4.13	1.4.10
HCatalog	2.3.6	2.3.6	2.3.6	2.3.6
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.6	2.3.6	2.3.6	2.3.6

	emr-5.30.2	emr-5.30.1	emr-5.30.0	emr-5.29.0
Hudi	0.5.2-incubating	0.5.2-incubating	0.5.2-incubating	0.5.0-incubating
Hue	4.6.0	4.6.0	4.6.0	4.4.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.1.0	1.1.0	1.1.0	1.0.0
Livy	0.7.0	0.7.0	0.7.0	0.6.0
MXNet	1.5.1	1.5.1	1.5.1	1.5.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.0	5.2.0	5.2.0	5.1.0
Phoenix	4.14.3	4.14.3	4.14.3	4.14.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.232	0.232	0.232	0.227
Spark	2.4.5	2.4.5	2.4.5	2.4.4
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.14.0	1.14.0	1.14.0	1.14.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.2	0.8.2	0.8.2	0.8.2
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

## 5.30.2 Versionshinweise

Dies ist eine Version zur Behebung von Problemen mit Amazon-EMR-Skalierung, wenn ein Cluster nicht erfolgreich hoch-/herunterskaliert werden kann oder Anwendungsfehler verursacht werden.

### Änderungen, Verbesserungen und behobene Probleme

- Es wurde ein Problem behoben, bei dem Skalierungsanforderungen für einen großen, stark ausgelasteten Cluster fehlschlagen, wenn Amazon-EMR-On-Cluster-Daemons Aktivitäten zur Integritätsprüfung durchführten, z. B. das Erfassen des YARN-Knotenstatus und des HDFS-Knotenstatus. Das lag daran, dass Cluster-Daemons die Integritätsstatusdaten eines Knotens nicht an interne Amazon-EMR-Komponenten weitergeben konnten.
- Verbesserte EMR-On-Cluster-Daemons zur korrekten Nachverfolgung der Knotenstatus bei der Wiederverwendung von IP-Adressen, um die Zuverlässigkeit bei Skalierungsvorgängen zu verbessern.
- [SPARK-29683](#). Es wurde ein Problem behoben, bei dem während der Cluster-Skalierung Auftragsfehler auftraten, da Spark davon ausging, dass alle verfügbaren Knoten auf der Verweigern-Liste standen.
- [YARN-9011](#). Es wurde ein Problem behoben, bei dem Auftragsfehler aufgrund eines Fehlers bei der Außerbetriebnahme von YARN auftraten, wenn der Cluster versuchte, hoch- oder herunterzuskalieren.
- Das Problem mit Schritt- oder Auftragsfehlern bei der Cluster-Skalierung wurde behoben, indem sichergestellt wurde, dass die Knotenstatus zwischen den Amazon-EMR-On-Cluster-Daemons und YARN/HDFS immer konsistent sind.
- Es wurde ein Problem behoben, bei dem Clustervorgänge wie Herunterskalierung und Schrittübermittlung für Amazon-EMR-Cluster, die mit Kerberos-Authentifizierung aktiviert waren, fehlschlagen. Dies lag daran, dass der Amazon-EMR-On-Cluster-Daemon das Kerberos-Ticket nicht erneuert hat, das für die sichere Kommunikation mit HDFS/YARN erforderlich ist, das auf dem Primärknoten ausgeführt wird.
- Neuere Amazon-EMR-Versionen beheben das Problem mit einem niedrigeren Limit für die „maximale Anzahl geöffneter Dateien“ auf älteren AL2-Versionen in Amazon EMR. Die Amazon-EMR-Versionen 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 und höher enthalten jetzt einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“.
- HTTPS ist jetzt standardmäßig für Amazon-Linux-Repositorys aktiviert. Wenn Sie eine Amazon-S3-VPCE-Richtlinie verwenden, um den Zugriff auf bestimmte Buckets zu beschränken, müssen Sie den neuen Amazon-Linux-Bucket-ARN `arn:aws:s3:::amazonlinux-2-repos-$region/*`



zu Ihrer Richtlinie hinzufügen (ersetzen Sie `$region` durch die Region, in der sich der Endpunkt befindet). Weitere Informationen finden Sie unter diesem Thema in den AWS Diskussionsforen. [Ankündigung: Amazon Linux 2 unterstützt jetzt die Möglichkeit, HTTPS zu verwenden, wenn eine Verbindung zu Paket-Repositorys hergestellt wird.](#)

## Bekannte Probleme

- Wenn Sie Spark mit der Formatierung des Hive-Partitionsstandorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den Amazon-EMR-Versionen 5.30.0 bis 5.36.0 und 6.2.0 bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:
  - Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.
  - Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist `s3://bucket/table/p=a` ein Präfix von `s3://bucket/table/p=a b`.
  - Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen UTF-8-Wert, der kleiner als das `/`-Zeichen (U+002F) ist. Beispielsweise fällt das Leerzeichen (U+0020), das in `s3://bucket/table/p=a b` zwischen `a` und `b` vorkommt, in diese Kategorie. Beachten Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: `!"#$%&'()*+,-.` Weitere Informationen finden Sie unter [UTF-8-Kodierungstabelle und Unicode-Zeichen](#).

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.

## 5.30.2 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-*amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-

Community-Komponente mit dem Namen myapp-component der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als 2.2-amzn-2 aufgeführt.

Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1.3.0	Amazon SageMaker Spark-SDK
emr-ddb	4.14.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.13.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.5.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.14.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	1.5.0	EMR S3Select-Konnektor
emrfs	2.40.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.10.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.

Komponente	Version	Beschreibung
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.5-amzn-6.1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.5-amzn-6.1	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.5-amzn-6.1	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.8.5-amzn-6.1	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-hdfs-journalnode	2.8.5-amzn-6.1	HDFS-Service zum Verwalten des Hadoop-Dateisystem journals auf HA-Clustern.
hadoop-httpfs-server	2.8.5-amzn-6.1	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.8.5-amzn-6.1	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider

Komponente	Version	Beschreibung
hadoop-mapred	2.8.5-amzn-6.1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.8.5-amzn-6.1	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.8.5-amzn-6.1	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.5-amzn-6.1	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.4.13	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.4.13	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.4.13	HBase-Befehlszeilen-Client.
hbase-rest-server	1.4.13	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.4.13	Service, der einen Thrift-Endpunkt für HBase bereitstellt.

Komponente	Version	Beschreibung
hcatalog-client	2.3.6-amzn-2	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.6-amzn-2	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.3.6-amzn-2	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.3.6-amzn-2	Hive-Befehlszeilen-Client.
hive-hbase	2.3.6-amzn-2	Hive-hbase client.
hive-metastore-server	2.3.6-amzn-2	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.3.6-amzn-2	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.5.2-incubating	Inkrementelles Verarbeitungs-Framework zur Stromversorgung der Datenpipeline bei geringer Latenz und hoher Effizienz.
hudi-presto	0.5.2-incubating	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.

Komponente	Version	Beschreibung
hue-server	4.6.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	1.1.0	Multi-User-Server für Jupyter-Notebooks
livy-server	0.7.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP- und Reverse-Proxy-Server.
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.5.1	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.64+	MySQL-Datenbankserver.
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Toolkit
oozie-client	5.2.0	Oozie-Befehlszeilen-Client.
oozie-server	5.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.3-HBase-1.4	Die Phoenix-Bibliotheken für den Server und den Client

Komponente	Version	Beschreibung
phoenix-query-server	4.14.3-HBase-1.4	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.232	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.232	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.232	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.3	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	1.2.0	Apache Ranger Key Management System
spark-client	2.4.5-amzn-0.1	Spark-Befehlszeilen-Clients.
spark-history-server	2.4.5-amzn-0.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.

Komponente	Version	Beschreibung
spark-on-yarn	2.4.5-amzn-0.1	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.4.5-amzn-0.1	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	1.14.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.2	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.8.2	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.14	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.14	ZooKeeper Befehlszeilen-Client.



## 5.30.2 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

### emr-5.30.2-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-log4j	Ändert die Werte in der <code>container-log4j.properties</code> -Datei in Hadoop YARN.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Ändert die <code>log4j-cli.properties</code> -Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.

Klassifizierungen	Beschreibung
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.

Klassifizierungen	Beschreibung
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.

Klassifizierungen	Beschreibung
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
hudi-env	Ändern der Werte in der Hudi-Umgebung.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.

Klassifizierungen	Beschreibung
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.

Klassifizierungen	Beschreibung
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.
ranger-kms-dbks-site	Ändert die Werte in der dbks-site.xml-Datei von Ranger KMS.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger KMS.
ranger-kms-env	Ändert die Werte in der Ranger KMS-Umgebung.
ranger-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei von Ranger KMS.
ranger-kms-db-ca	Ändert die Werte für die CA-Datei auf S3 für die MySQL SSL-Verbindung mit Ranger KMS.
recordserver-env	Werte in der RecordServer EMR-Umgebung ändern.
recordserver-conf	Ändern Sie die Werte in der Datei RecordServer server.properties von EMR.
recordserver-log4j	Ändern Sie die Werte in der Datei RecordServer log4j.properties von EMR.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.

Klassifizierungen	Beschreibung
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml von Sqoop OraOop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.



## Amazon-EMR-Version 5.30.1

### 5.30.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

#### Informationen zur Anwendungsversion

	emr-5.30.1	emr-5.30.0	emr-5.29.0	emr-5.28.1
AWS SDK for Java	1.11.759	1,11.759	1,11,682	1,11,659
Python	2,7, 3,7	2,7, 3,7	2,7, 3,6	2,7, 3,6
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.10.0	1.10.0	1.9.1	1.9.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2

	emr-5.30.1	emr-5.30.0	emr-5.29.0	emr-5.28.1
HBase	1.4.13	1.4.13	1.4.10	1.4.10
HCatalog	2.3.6	2.3.6	2.3.6	2.3.6
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.6	2.3.6	2.3.6	2.3.6
Hudi	0.5.2-incubating	0.5.2-incubating	0.5.0-incubating	0.5.0-incubating
Hue	4.6.0	4.6.0	4.4.0	4.4.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.1.0	1.1.0	1.0.0	1.0.0
Livy	0.7.0	0.7.0	0.6.0	0.6.0
MXNet	1.5.1	1.5.1	1.5.1	1.5.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.0	5.2.0	5.1.0	5.1.0
Phoenix	4.14.3	4.14.3	4.14.3	4.14.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.232	0.232	0.227	0.227
Spark	2.4.5	2.4.5	2.4.4	2.4.4
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.14.0	1.14.0	1.14.0	1.14.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2

	emr-5.30.1	emr-5.30.0	emr-5.29.0	emr-5.28.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.2	0.8.2	0.8.2	0.8.2
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

## 5.30.1 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.30.1. Änderungen beziehen sich auf Version 5.30.0.

Erstveröffentlichungsdatum: 30. Juni 2020

Letzte Aktualisierung: 24. August 2020

### Änderungen, Verbesserungen und behobene Probleme

- Neuere Amazon-EMR-Versionen beheben das Problem mit einem niedrigeren Limit für die „maximale Anzahl geöffneter Dateien“ auf älteren AL2-Versionen in Amazon EMR. Die Amazon-EMR-Versionen 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 und höher enthalten jetzt einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“.
- Es wurde ein Problem behoben, bei dem der Instance-Controller-Prozess unendlich viele Prozesse ausgelöst hat.
- Es wurde ein Problem behoben, bei dem Hue keine Hive-Abfrage ausführen konnte, die Meldung „Datenbank ist gesperrt“ angezeigt wurde und die Ausführung von Abfragen verhindert wurde.
- Es wurde ein Spark-Problem behoben, durch das mehr Aufgaben gleichzeitig auf dem EMR-Cluster ausgeführt werden konnten.
- Es wurde ein Problem mit dem Jupyter Notebook behoben, das den Fehler „Zu viele Dateien öffnen“ auf dem Jupyter-Server verursachte.
- Ein Problem mit den Startzeiten von Clustern wurde behoben.

## Neue Features

- Persistente Anwendungsschnittstellen für die Tez-Benutzeroberfläche und den YARN-Timeline-Server sind mit den Amazon-EMR-Versionen 6.x und EMR-Version 5.30.1 und höher verfügbar. Durch den Linkzugriff mit einem Klick auf den persistenten Anwendungsverlauf können Sie schnell auf den Jobverlauf zugreifen, ohne einen Webproxy über eine SSH-Verbindung einrichten zu müssen. Protokolle für aktive und beendete Cluster sind 30 Tage nach Ende der Anwendung verfügbar. Weitere Informationen finden Sie unter [Ansicht der Benutzeroberflächen für persistente Anwendungen](#) im Verwaltungshandbuch für Amazon EMR an.
- EMR-Notebook-Ausführungs-APIs sind verfügbar, um EMR-Notebooks über ein Skript oder eine Befehlszeile auszuführen. Durch die Möglichkeit, EMR-Notebookausführungen ohne die AWS Konsole zu starten, zu stoppen, aufzulisten und zu beschreiben, können Sie ein EMR-Notebook programmgesteuert steuern. Mithilfe einer parametrisierten Notebook-Zelle können Sie verschiedene Parameterwerte an ein Notebook übergeben, ohne für jeden neuen Satz von Parameterwerten eine Kopie des Notebooks erstellen zu müssen. Siehe [EMR-API-Aktionen](#). Beispielcode finden Sie unter [Beispielbefehle zur programmatischen Ausführung von EMR Notebooks](#).

## Bekannte Probleme

- Niedrigeres Limit für die „Max. Anzahl geöffneter Dateien“ auf älteren AL2-Versionen [in neueren Versionen behoben]. Amazon-EMR-Versionen: emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0 und emr-6.2.0 basieren auf älteren Versionen von Amazon Linux 2 (AL2), die eine niedrigere Ulimit-Einstellung für „Max. Anzahl geöffneter Dateien“ haben, wenn Amazon-EMR-Cluster mit dem Standard-AMI erstellt werden. Die Amazon-EMR-Versionen 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 und höher enthalten einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“. Versionen mit einem niedrigeren Limit für geöffnete Dateien verursachen beim Senden des Spark-Jobs den Fehler „Zu viele offene Dateien“. In den betroffenen Versionen hat das Amazon-EMR-Standard-AMI eine Ulimit-StandardEinstellung von 4 096 für „Max. Anzahl geöffneter Dateien“, was unter dem Dateilimit von 65 536 im neuesten Amazon-Linux-2-AMI liegt. Die niedrigere Ulimit-Einstellung für „Max. Anzahl geöffneter Dateien“ führt dazu, dass der Spark-Job fehlschlägt, wenn der Spark-Treiber und der Executor versuchen, mehr als 4 096 Dateien zu öffnen. Um das Problem zu beheben, verfügt Amazon EMR über ein Bootstrap Action (BA)-Skript, das die Ulimit-Einstellung bei der Cluster-Erstellung anpasst.

Wenn Sie eine ältere Amazon-EMR-Version verwenden, für die dieses Problem nicht dauerhaft behoben ist, können Sie mit der folgenden Problemumgehung das Instance-Controller-Ulimit explizit auf maximal 65 536 Dateien festlegen.

Explizit ein ulimit über die Befehlszeile setzen

1. Bearbeiten Sie `/etc/systemd/system/instance-controller.service`, um die folgenden Parameter zum Abschnitt Service hinzuzufügen.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Starten Sie neu InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Mithilfe der Bootstrap-Aktion (BA) ein Ulimit festlegen

Sie können auch ein Bootstrap-Aktionsskript (BA) verwenden, um das Ulimit für den Instance-Controller bei der Clustererstellung auf 65 536 Dateien zu konfigurieren.

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
```

```
sudo systemctl daemon-reload
```

- EMR Notebooks

Das Feature, mit der Sie Kernel und zusätzliche Python-Bibliotheken auf dem Cluster-Primärknoten installieren können, ist in EMR-Version 5.30.1 standardmäßig deaktiviert. Weitere Informationen zu diesem Feature finden Sie unter [Installation von Kernels und Python-Bibliotheken auf einem Cluster-Primärknoten](#).

Um das Feature zu aktivieren, gehen Sie wie folgt vor:

1. Stellen Sie sicher, dass die der Servicerolle für EMR Notebooks zugeordnete Berechtigungsrichtlinie die folgende Aktion zulässt:

```
elasticmapreduce:ListSteps
```

Weitere Informationen finden Sie unter [Servicerolle für EMR-Notebooks](#).

2. Verwenden Sie den AWS CLI , um einen Schritt auf dem Cluster auszuführen, der EMR Notebooks einrichtet, wie im folgenden Beispiel gezeigt. Ersetzen Sie *us-east-1* durch die Region, in der sich Ihr Cluster befindet. Weitere Informationen finden Sie unter [Hinzufügen von Schritten zu einem Cluster mithilfe AWS CLI](#).

```
aws emr add-steps --cluster-id MyClusterID --steps
  Type=CUSTOM_JAR,Name=EMRNotebooksSetup,ActionOnFailure=CONTINUE,Jar=s3://us-
east-1.elasticmapreduce/libs/script-runner/script-runner.jar,Args=["s3://
awssupportdatasvcs.com/bootstrap-actions/EMRNotebooksSetup/emr-notebooks-setup.sh"]
```

- Verwaltete Skalierung

Verwaltete Skalierungsvorgänge auf Clustern der Versionen 5.30.0 und 5.30.1, ohne dass Presto installiert ist, können zu Anwendungsausfällen führen oder dazu führen, dass eine einheitliche Instance-Gruppe oder Instance-Flotte unverändert im Status ARRESTED bleibt, insbesondere wenn auf einen Herunterskalierungsvorgang schnell ein Skalierungsvorgang folgt.

Um dieses Problem zu umgehen, wählen Sie Presto als zu installierende Anwendung, wenn Sie einen Cluster mit den Amazon-EMR-Versionen 5.30.0 und 5.30.1 erstellen, auch wenn Ihr Auftrag Presto nicht benötigt.

- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in Amazon-EMR-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie dem Herunterskalieren oder der schrittweisen Übermittlung auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

Workaround:

- SSH als hadoop-Benutzer für den führenden Primärknoten des EMR-Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den hadoop-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

#### Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

- Wenn Sie Spark mit der Formatierung des Hive-Partitionsstandorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den Amazon-EMR-Versionen 5.30.0 bis 5.36.0 und 6.2.0 bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:
  - Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.
  - Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist `s3://bucket/table/p=a` ein Präfix von `s3://bucket/table/p=a b`.
  - Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen UTF-8-Wert, der kleiner als das `/`-Zeichen (U+002F) ist. Beispielsweise fällt das Leerzeichen (U+0020),

das in `s3://bucket/table/p=a b` zwischen `a` und `b` vorkommt, in diese Kategorie. Beachten Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: `!"#$%&'()*+,-.` Weitere Informationen finden Sie unter [UTF-8-Kodierungstabelle und Unicode-Zeichen](#).

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.

### 5.30.1 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.3.0	Amazon SageMaker Spark-SDK
<code>emr-ddb</code>	4.14.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.13.0	Praktische Bibliotheken für das Hadoop-Ökosystem.



Komponente	Version	Beschreibung
emr-kinesis	3.5.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.14.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	1.5.0	EMR S3Select-Konnektor
emrfs	2.40.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.10.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.5-amzn-6	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".

Komponente	Version	Beschreibung
hadoop-hdfs-datanode	2.8.5-amzn-6	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.5-amzn-6	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.8.5-amzn-6	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-hdfs-journalnode	2.8.5-amzn-6	HDFS-Service zum Verwalten des Hadoop-Dateisystem journals auf HA-Clustern.
hadoop-httfs-server	2.8.5-amzn-6	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.8.5-amzn-6	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.8.5-amzn-6	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.8.5-amzn-6	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.8.5-amzn-6	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.

Komponente	Version	Beschreibung
hadoop-yarn-timeline-server	2.8.5-amzn-6	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.4.13	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.4.13	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.4.13	HBase-Befehlszeilen-Client.
hbase-rest-server	1.4.13	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.4.13	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.6-amzn-2	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.6-amzn-2	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.3.6-amzn-2	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.

Komponente	Version	Beschreibung
hive-client	2.3.6-amzn-2	Hive-Befehlszeilen-Client.
hive-hbase	2.3.6-amzn-2	Hive-hbase client.
hive-metastore-server	2.3.6-amzn-2	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.3.6-amzn-2	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.5.2-incubating	Inkrementelles Verarbeitungs-Framework zur Stromversorgung der Datenpipeline bei geringer Latenz und hoher Effizienz.
hudi-presto	0.5.2-incubating	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hue-server	4.6.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	1.1.0	Multi-User-Server für Jupyter-Notebooks
livy-server	0.7.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP- und Reverse-Proxy-Server.

Komponente	Version	Beschreibung
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.5.1	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.64	MySQL-Datenbankserver.
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.2.0	Oozie-Befehlszeilen-Client.
oozie-server	5.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.3-HBase-1.4	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.14.3-HBase-1.4	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.232	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.232	Service für das Ausführen von Teilen einer Abfrage.

Komponente	Version	Beschreibung
presto-client	0.232	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.3	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	1.2.0	Apache Ranger Key Management System
spark-client	2.4.5-amzn-0	Spark-Befehlszeilen-Clients.
spark-history-server	2.4.5-amzn-0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.4.5-amzn-0	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.4.5-amzn-0	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.

Komponente	Version	Beschreibung
tensorflow	1.14.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.2	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.8.2	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.14	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.14	ZooKeeper Befehlszeilen-Cli-ent.

### 5.30.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

## emr-5.30.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der capacity-scheduler.xml-Datei in Hadoop.
container-log4j	Ändert die Werte in der container-log4j.properties-Datei in Hadoop YARN.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4j-yarn-session.properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.



Klassifizierungen	Beschreibung
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.

Klassifizierungen	Beschreibung
hive-parquet-logging	Ändert die Werte in der <code>parquet-logging.properties</code> -Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von <code>hive-exec-log</code> Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von Hive. <code>llap-daemon-log</code>
hive-log4j2	Ändert die Werte in der <code>hive-log4j2.properties</code> -Datei in Hive.
hive-site	Ändert die Werte in der <code>hive-site.xml</code> -Datei in Hive.
hiveserver2-site	Ändert die Werte in der <code>hiveserver2-site.xml</code> -Datei von <code>Server2</code> in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der <code>httpfs-site.xml</code> -Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der <code>kms-acls.xml</code> -Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der <code>kms-log4j.properties</code> -Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der <code>kms-site.xml</code> -Datei in Hadoop.

Klassifizierungen	Beschreibung
hudi-env	Ändern der Werte in der Hudi-Umgebung.
jupyter-notebook-conf	Ändert die Werte in der <code>jupyter_notebook_config.py</code> -Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei <code>jupyterhub_config.py</code> .
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der <code>config.json</code> -Datei in Sparkmagic.
livy-conf	Ändert die Werte in der <code>livy.conf</code> -Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei <code>mapred-site.xml</code> der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der <code>oozie-log4j.properties</code> -Datei in Oozie.
oozie-site	Ändert die Werte in der <code>oozie-site.xml</code> -Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der <code>hadoop-metrics2-hbase.properties</code> -Datei in Phoenix.

Klassifizierungen	Beschreibung
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.

Klassifizierungen	Beschreibung
ranger-kms-dbks-site	Ändert die Werte in der dbks-site.xml-Datei von Ranger KMS.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger KMS.
ranger-kms-env	Ändert die Werte in der Ranger KMS-Umgebung.
ranger-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei von Ranger KMS.
ranger-kms-db-ca	Ändert die Werte für die CA-Datei auf S3 für die MySQL SSL-Verbindung mit Ranger KMS.
recordserver-env	Werte in der RecordServer EMR-Umgebung ändern.
recordserver-conf	Ändern Sie die Werte in der Datei RecordServer server.properties von EMR.
recordserver-log4j	Ändern Sie die Werte in der Datei RecordServer log4j.properties von EMR.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.

Klassifizierungen	Beschreibung
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml von Sqoop OraOop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.30.0

### 5.30.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

#### Informationen zur Anwendungsversion

	emr-5.30.0	emr-5.29.0	emr-5.28.1	emr-5.28.0
AWS SDK for Java	1.11.759	1,11,682	1,11,659	1,11,659
Python	2,7, 3,7	2,7, 3,6	2,7, 3,6	2,7, 3,6
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.10.0	1.9.1	1.9.0	1.9.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.13	1.4.10	1.4.10	1.4.10
HCatalog	2.3.6	2.3.6	2.3.6	2.3.6
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.6	2.3.6	2.3.6	2.3.6



	emr-5.30.0	emr-5.29.0	emr-5.28.1	emr-5.28.0
Hudi	0.5.2-incubating	0.5.0-incubating	0.5.0-incubating	0.5.0-incubating
Hue	4.6.0	4.4.0	4.4.0	4.4.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.1.0	1.0.0	1.0.0	1.0.0
Livy	0.7.0	0.6.0	0.6.0	0.6.0
MXNet	1.5.1	1.5.1	1.5.1	1.5.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.0	5.1.0	5.1.0	5.1.0
Phoenix	4.14.3	4.14.3	4.14.3	4.14.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.232	0.227	0.227	0.227
Spark	2.4.5	2.4.4	2.4.4	2.4.4
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.14.0	1.14.0	1.14.0	1.14.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.2	0.8.2	0.8.2	0.8.2
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

## 5.30.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.30.0. Änderungen beziehen sich auf Version 5.29.0.

Erste Version: 13. Mai 2020

Letzte Aktualisierung: 25. Juni 2020

### Upgrades

- Auf Version AWS SDK for Java 1.11.759 aktualisiert
- Amazon SageMaker Spark SDK wurde auf Version 1.3.0 aktualisiert
- EMR Record Server auf Version 1.6.0 aktualisiert
- Flink auf Version 1.10.0 aktualisiert
- Ganglia auf Version 3.7.2 aktualisiert
- HBase auf Version 1.4.13 aktualisiert
- Hudi auf Version 0.5.2-incubating aktualisiert
- Hue auf Version 4.6.0 aktualisiert
- Auf Version JupyterHub 1.1.0 aktualisiert
- Livy auf Version 0.7.0-incubating aktualisiert
- Oozie auf Version 5.2.0 aktualisiert
- Presto auf Version 0.232 aktualisiert
- Spark auf Version 2.4.5 aktualisiert
- Konnektoren und Treiber aktualisiert: Amazon Glue Connector 1.12.0; Amazon Kinesis Connector 3.5.0; EMR DynamoDB Connector 4.14.0

### Neue Features

- EMR Notebooks – Bei Verwendung mit EMR-Clustern, die mit 5.30.0 erstellt wurden, werden EMR-Notebooks-Kernel auf dem Cluster ausgeführt. Dies verbessert die Notebook-Leistung und ermöglicht es Ihnen, Kernel zu installieren und anzupassen. Sie können Python-Bibliotheken auch auf dem Cluster-Primärknoten installieren. Weitere Informationen finden Sie unter [Installieren und Verwenden von Kernels und Bibliotheken](#) im Management Guide für EMR.

- **Verwaltete Skalierung** – Mit Amazon EMR ab Version 5.30.0 können Sie die verwaltete EMR-Skalierung aktivieren, um die Anzahl der Instances oder Einheiten in Ihrem Cluster basierend auf der Workload automatisch zu erhöhen oder zu verringern. Amazon EMR wertet Cluster-Metriken kontinuierlich aus, um Skalierungsentscheidungen zu treffen, die Ihre Cluster für Kosten und Geschwindigkeit optimieren. Weitere Informationen finden Sie unter [Skalieren von Clusterressourcen](#) im Verwaltungshandbuch für Amazon EMR.
- **In Amazon S3 gespeicherte Protokolldateien verschlüsseln** — Mit Amazon EMR Version 5.30.0 und höher können Sie in Amazon S3 gespeicherte Protokolldateien mit einem vom Kunden verwalteten Schlüssel verschlüsseln. AWS KMS Weitere Informationen finden Sie unter [Verschlüsselte gespeicherte Protokolldateien in Amazon S3](#) im Verwaltungshandbuch für Amazon EMR.
- **Unterstützung von Amazon Linux 2** – In EMR Version 5.30.0 und höher verwendet EMR das Betriebssystem Amazon Linux 2. Neue benutzerdefinierte AMIs (Amazon Machine Image) müssen auf dem Amazon-Linux-2-AMI basieren. Weitere Informationen finden Sie unter [Verwenden eines benutzerdefinierten AMI](#).
- **Ordnungsgemäßes Presto Auto Scaling** – EMR-Cluster mit 5.30.0 können mit einer Auto-Scaling-Zeitüberschreitung festgelegt werden, die Presto-Aufgaben Zeit zum Abschluss der Ausführung gibt, bevor ihr Knoten stillgelegt wird. Weitere Informationen finden Sie unter [Auto Scaling von Presto mit ordnungsgemäßer Stilllegung verwenden](#).
- **Flotteninstance-Erstellung mit neuer Zuweisungsstrategieoption** – Eine neue Zuweisungsstrategieoption ist in EMR-Version 5.12.1 und höher verfügbar. Sie bietet eine schnellere Cluster-Bereitstellung, eine genauere Spot-Zuweisung und weniger Unterbrechungen von Spot Instances. Aktualisierungen für nicht standardmäßige EMR-Service rollen sind erforderlich. Sehen Sie unter [Konfigurieren von Instance-Flotten](#).
- **Befehle `sudo systemctl stop` und `sudo systemctl start`** – In EMR Version 5.30.0 und höher, die das Betriebssystem Amazon Linux 2 verwenden, verwendet EMR die Befehle `sudo systemctl stop` und `sudo systemctl start`, um Services neu zu starten. Weitere Informationen finden Sie unter [Wie starte ich einen Service in Amazon EMR?](#).

## Änderungen, Verbesserungen und behobene Probleme

- EMR Version 5.30.0 installiert Ganglia standardmäßig nicht. Sie können Ganglia explizit für die Installation auswählen, wenn Sie einen Cluster erstellen.
- Spark-Leistungsoptimierungen.
- Presto-Leistungsoptimierungen.

- Python 3 ist der Standard für Amazon-EMR-Version 5.30.0 und höher.
- Die standardmäßige verwaltete Sicherheitsgruppe für den Zugriff auf Services in privaten Subnetzen wurde mit neuen Regeln aktualisiert. Wenn Sie benutzerdefinierte Sicherheitsgruppe für den Servicezugriff verwenden, müssen Sie dieselben Regeln wie die standardmäßige verwaltete Sicherheitsgruppe einschließen. Weitere Informationen finden Sie unter [Amazon-EMR-verwaltete Sicherheitsgruppe für den Servicezugriff \(private Subnetze\)](#). Wenn Sie eine benutzerdefinierte Servicerolle für Amazon EMR verwenden, müssen Sie die Berechtigung zum `ec2:describeSecurityGroups` erteilen, damit EMR überprüfen kann, ob die Sicherheitsgruppen korrekt erstellt wurden. Wenn Sie `EMR_DefaultRole` verwenden, ist diese Berechtigung bereits in der standardmäßigen verwalteten Richtlinie enthalten.

## Bekannte Probleme

- Niedrigeres Limit für die „Max. Anzahl geöffneter Dateien“ auf älteren AL2-Versionen [in neueren Versionen behoben]. Amazon-EMR-Versionen: `emr-5.30.x`, `emr-5.31.0`, `emr-5.32.0`, `emr-6.0.0`, `emr-6.1.0` und `emr-6.2.0` basieren auf älteren Versionen von Amazon Linux 2 (AL2), die eine niedrigere `Ulimit`-Einstellung für „Max. Anzahl geöffneter Dateien“ haben, wenn Amazon-EMR-Cluster mit dem Standard-AMI erstellt werden. Die Amazon-EMR-Versionen 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 und höher enthalten einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“. Versionen mit einem niedrigeren Limit für geöffnete Dateien verursachen beim Senden des Spark-Jobs den Fehler „Zu viele offene Dateien“. In den betroffenen Versionen hat das Amazon-EMR-Standard-AMI eine `Ulimit`-Standardeinstellung von 4 096 für „Max. Anzahl geöffneter Dateien“, was unter dem Dateilimit von 65 536 im neuesten Amazon-Linux-2-AMI liegt. Die niedrigere `Ulimit`-Einstellung für „Max. Anzahl geöffneter Dateien“ führt dazu, dass der Spark-Job fehlschlägt, wenn der Spark-Treiber und der Executor versuchen, mehr als 4 096 Dateien zu öffnen. Um das Problem zu beheben, verfügt Amazon EMR über ein Bootstrap Action (BA)-Skript, das die `Ulimit`-Einstellung bei der Cluster-Erstellung anpasst.

Wenn Sie eine ältere Amazon-EMR-Version verwenden, für die dieses Problem nicht dauerhaft behoben ist, können Sie mit der folgenden Problemumgehung das Instance-Controller-`Ulimit` explizit auf maximal 65 536 Dateien festlegen.

Explizit ein `ulimit` über die Befehlszeile setzen

1. Bearbeiten Sie `/etc/systemd/system/instance-controller.service`, um die folgenden Parameter zum Abschnitt Service hinzuzufügen.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

## 2. Starten Sie neu InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

### Mithilfe der Bootstrap-Aktion (BA) ein Ulimit festlegen

Sie können auch ein Bootstrap-Aktionsskript (BA) verwenden, um das Ulimit für den Instance-Controller bei der Clustererstellung auf 65 536 Dateien zu konfigurieren.

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

- **Verwaltete Skalierung**

Verwaltete Skalierungsvorgänge auf Clustern der Versionen 5.30.0 und 5.30.1, ohne dass Presto installiert ist, können zu Anwendungsausfällen führen oder dazu führen, dass eine einheitliche Instance-Gruppe oder Instance-Flotte unverändert im Status ARRESTED bleibt, insbesondere wenn auf einen Herunterskalierungsvorgang schnell ein Skalierungsvorgang folgt.

Um dieses Problem zu umgehen, wählen Sie Presto als zu installierende Anwendung, wenn Sie einen Cluster mit den Amazon-EMR-Versionen 5.30.0 und 5.30.1 erstellen, auch wenn Ihr Auftrag Presto nicht benötigt.

- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in Amazon-EMR-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie dem Herunterskalieren oder der schrittweisen Übermittlung auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

Workaround:

- SSH als hadoop-Benutzer für den führenden Primärknoten des EMR-Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den hadoop-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

#### Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

- Die Standard-Datenbank-Engine für Hue 4.6.0 ist SQLite, was zu Problemen führt, wenn Sie versuchen, Hue mit einer externen Datenbank zu verwenden. Um dieses Problem zu beheben, setzen Sie `engine` in Ihrer `hue.ini` Konfigurationsklassifizierung auf `mysql`. Dieses Problem wurde in Amazon-EMR-Version 5.30.1 behoben.

- Wenn Sie Spark mit der Formatierung des Hive-Partitionsstandorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den Amazon-EMR-Versionen 5.30.0 bis 5.36.0 und 6.2.0 bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:
  - Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.
  - Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist `s3://bucket/table/p=a` ein Präfix von `s3://bucket/table/p=a b`.
  - Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen UTF-8-Wert, der kleiner als das `/`-Zeichen (U+002F) ist. Beispielsweise fällt das Leerzeichen (U+0020), das in `s3://bucket/table/p=a b` zwischen `a` und `b` vorkommt, in diese Kategorie. Beachten Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: `!"#$%&'()*+,-.` Weitere Informationen finden Sie unter [UTF-8-Kodierungstabelle und Unicode-Zeichen](#).

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.

## 5.30.0 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-`amzn-EmrVersion`. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1.3.0	Amazon SageMaker Spark-SDK
emr-ddb	4.14.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.13.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.5.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-notebook-env	1.0.0	Conda env für emr Notebook
emr-s3-dist-cp	2.14.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	1.5.0	EMR S3Select-Konnektor
emrfs	2.40.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.10.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.



Komponente	Version	Beschreibung
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.5-amzn-6	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.5-amzn-6	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.5-amzn-6	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.8.5-amzn-6	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-hdfs-journalnode	2.8.5-amzn-6	HDFS-Service zum Verwalten des Hadoop-Dateisystem journals auf HA-Clustern.
hadoop-httpfs-server	2.8.5-amzn-6	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.8.5-amzn-6	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider

Komponente	Version	Beschreibung
hadoop-mapred	2.8.5-amzn-6	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.8.5-amzn-6	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.8.5-amzn-6	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.5-amzn-6	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.4.13	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.4.13	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.4.13	HBase-Befehlszeilen-Client.
hbase-rest-server	1.4.13	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.4.13	Service, der einen Thrift-Endpunkt für HBase bereitstellt.

Komponente	Version	Beschreibung
hcatalog-client	2.3.6-amzn-2	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.6-amzn-2	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.3.6-amzn-2	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.3.6-amzn-2	Hive-Befehlszeilen-Client.
hive-hbase	2.3.6-amzn-2	Hive-hbase client.
hive-metastore-server	2.3.6-amzn-2	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.3.6-amzn-2	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.5.2-incubating	Inkrementelles Verarbeitungs-Framework zur Stromversorgung der Datenpipeline bei geringer Latenz und hoher Effizienz.
hudi-presto	0.5.2-incubating	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.

Komponente	Version	Beschreibung
hue-server	4.6.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	1.1.0	Multi-User-Server für Jupyter-Notebooks
livy-server	0.7.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP- und Reverse-Proxy-Server.
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.5.1	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mariadb-server	5.5.64	MySQL-Datenbankserver.
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Toolkit
oozie-client	5.2.0	Oozie-Befehlszeilen-Client.
oozie-server	5.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.3-HBase-1.4	Die Phoenix-Bibliotheken für den Server und den Client

Komponente	Version	Beschreibung
phoenix-query-server	4.14.3-HBase-1.4	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.232	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.232	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.232	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.3	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
ranger-kms-server	1.2.0	Apache Ranger Key Management System
spark-client	2.4.5-amzn-0	Spark-Befehlszeilen-Clients.
spark-history-server	2.4.5-amzn-0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.

Komponente	Version	Beschreibung
spark-on-yarn	2.4.5-amzn-0	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.4.5-amzn-0	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	1.14.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.2	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.8.2	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.14	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.14	ZooKeeper Befehlszeilen-Client.

## 5.30.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

### emr-5.30.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-log4j	Ändert die Werte in der <code>container-log4j.properties</code> -Datei in Hadoop YARN.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Ändert die <code>log4j-cli.properties</code> -Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.

Klassifizierungen	Beschreibung
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.



Klassifizierungen	Beschreibung
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.

Klassifizierungen	Beschreibung
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
hudi-env	Ändern der Werte in der Hudi-Umgebung.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.

Klassifizierungen	Beschreibung
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.

Klassifizierungen	Beschreibung
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.
ranger-kms-dbks-site	Ändert die Werte in der dbks-site.xml-Datei von Ranger KMS.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger KMS.
ranger-kms-env	Ändert die Werte in der Ranger KMS-Umgebung.
ranger-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei von Ranger KMS.
ranger-kms-db-ca	Ändert die Werte für die CA-Datei auf S3 für die MySQL SSL-Verbindung mit Ranger KMS.
recordserver-env	Werte in der RecordServer EMR-Umgebung ändern.
recordserver-conf	Ändern Sie die Werte in der Datei RecordServer server.properties von EMR.
recordserver-log4j	Ändern Sie die Werte in der Datei RecordServer log4j.properties von EMR.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.

Klassifizierungen	Beschreibung
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml von Sqoop OraOop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.29.0

### 5.29.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

#### Informationen zur Anwendungsversion

	emr-5.29.0	emr-5.28.1	emr-5.28.0	emr-5.27.1
AWS SDK for Java	1.11.682	1,11,659	1,11,659	1,11,615
Python	2,7, 3,6	2,7, 3,6	2,7, 3,6	2,7, 3,6
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.9.1	1.9.0	1.9.0	1.8.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2

	emr-5.29.0	emr-5.28.1	emr-5.28.0	emr-5.27.1
HBase	1.4.10	1.4.10	1.4.10	1.4.10
HCatalog	2.3.6	2.3.6	2.3.6	2.3.5
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.6	2.3.6	2.3.6	2.3.5
Hudi	0.5.0-incubating	0.5.0-incubating	0.5.0-incubating	-
Hue	4.4.0	4.4.0	4.4.0	4.4.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.0.0	1.0.0	1.0.0	1.0.0
Livy	0.6.0	0.6.0	0.6.0	0.6.0
MXNet	1.5.1	1.5.1	1.5.1	1.4.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.1.0	5.1.0	5.1.0	5.1.0
Phoenix	4.14.3	4.14.3	4.14.3	4.14.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.227	0.227	0.227	0.224
Spark	2.4.4	2.4.4	2.4.4	2.4.4
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.14.0	1.14.0	1.14.0	1.14.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2



	emr-5.29.0	emr-5.28.1	emr-5.28.0	emr-5.27.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.2	0.8.2	0.8.2	0.8.1
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

## 5.29.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.29.0. Änderungen beziehen sich auf Version 5.28.1.

Erste Version: 17. Januar 2020

### Upgrades

- Auf Version AWS SDK for Java 1.11.682 aktualisiert
- Upgrade von Hive auf Version 2.3.6
- Flink wurde auf Version 1.9.1 aktualisiert
- EmrFS wurde auf Version 2.38.0 aktualisiert
- EMR DynamoDB Connector wurde auf Version 4.13.0 aktualisiert

### Änderungen, Verbesserungen und behobene Probleme

- Spark
  - Spark-Leistungsoptimierungen.
- EMRFS
  - Management Guide aktualisiert die Standardeinstellungen von emrfs-site.xml für eine konsistente Ansicht.

### Bekannte Probleme

- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in Amazon-EMR-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie dem Herunterskalieren oder der schrittweisen Übermittlung auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

Workaround:

- SSH als hadoop-Benutzer für den führenden Primärknoten des EMR-Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den hadoop-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

#### Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

## 5.29.0 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-amzn-*EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen myapp-component der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als 2.2-amzn-2 aufgeführt.

Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1.2.6	Amazon SageMaker Spark-SDK
emr-ddb	4.13.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.12.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.13.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	1.4.0	EMR S3Select-Konnektor
emrfs	2.38.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.9.1	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus

Komponente	Version	Beschreibung
		dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.5-amzn-5	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.5-amzn-5	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.5-amzn-5	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.8.5-amzn-5	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-hdfs-journalnode	2.8.5-amzn-5	HDFS-Service zum Verwalten des Hadoop-Dateisystem journals auf HA-Clustern.
hadoop-httpfs-server	2.8.5-amzn-5	HTTP-Endpunkt für HDFS-Operationen.

Komponente	Version	Beschreibung
hadoop-kms-server	2.8.5-amzn-5	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.8.5-amzn-5	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.8.5-amzn-5	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.8.5-amzn-5	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.5-amzn-5	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.4.10	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.4.10	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.4.10	HBase-Befehlszeilen-Client.

Komponente	Version	Beschreibung
hbase-rest-server	1.4.10	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.4.10	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.6-amzn-1	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.6-amzn-1	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.3.6-amzn-1	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.3.6-amzn-1	Hive-Befehlszeilen-Client.
hive-hbase	2.3.6-amzn-1	Hive-hbase client.
hive-metastore-server	2.3.6-amzn-1	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.3.6-amzn-1	Service zur Annahme von Hive-Abfragen als Webanfragen.

Komponente	Version	Beschreibung
hudi	0.5.0-incubating	Inkrementelles Verarbeitungs-Framework zur Stromversorgung der Datenpipeline bei geringer Latenz und hoher Effizienz.
hudi-presto	0.5.0-incubating	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hue-server	4.4.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	1.0.0	Multi-User-Server für Jupyter-Notebooks
livy-server	0.6.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP- und Reverse-Proxy-Server.
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.5.1	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Toolkit
oozie-client	5.1.0	Oozie-Befehlszeilen-Client.

Komponente	Version	Beschreibung
oozie-server	5.1.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.3-HBase-1.4	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.14.3-HBase-1.4	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.227	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.227	Service für das Ausführen von Teilen einer Abfrage.
presto-client	0.227	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.1	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)



Komponente	Version	Beschreibung
spark-client	2.4.4	Spark-Befehlszeilen-Clients.
spark-history-server	2.4.4	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.4.4	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.4.4	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	1.14.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.2	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.8.2	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.

Komponente	Version	Beschreibung
zookeeper-server	3.4.14	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.14	ZooKeeper Befehlszeilen-Cli-ent.

## 5.29.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

### emr-5.29.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-log4j	Ändert die Werte in der <code>container-log4j.properties</code> -Datei in Hadoop YARN.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.

Klassifizierungen	Beschreibung
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4 j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.

Klassifizierungen	Beschreibung
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.

Klassifizierungen	Beschreibung
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.

Klassifizierungen	Beschreibung
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.

Klassifizierungen	Beschreibung
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.
ranger-kms-dbks-site	Ändert die Werte in der dbks-site.xml-Datei von Ranger KMS.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger KMS.
ranger-kms-env	Ändert die Werte in der Ranger KMS-Umgebung.
ranger-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei von Ranger KMS.
ranger-kms-db-ca	Ändert die Werte für die CA-Datei auf S3 für die MySQL SSL-Verbindung mit Ranger KMS.
recordserver-env	Werte in der RecordServer EMR-Umgebung ändern.



Klassifizierungen	Beschreibung
recordserver-conf	Ändern Sie die Werte in der Datei RecordServer server.properties von EMR.
recordserver-log4j	Ändern Sie die Werte in der Datei RecordServer log4j.properties von EMR.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml von Sqoop OraOop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.

Klassifizierungen	Beschreibung
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.28.1

### 5.28.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

## Informationen zur Anwendungsversion

	emr-5.28.1	emr-5.28.0	emr-5.27.1	emr-5.27.0
AWS SDK for Java	1.11.659	1,11.659	1,11,615	1,11,615
Python	2,7, 3,6	2,7, 3,6	2,7, 3,6	2,7, 3,6
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.9.0	1.9.0	1.8.1	1.8.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.10	1.4.10	1.4.10	1.4.10
HCatalog	2.3.6	2.3.6	2.3.5	2.3.5
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.6	2.3.6	2.3.5	2.3.5
Hudi	0.5.0-incubating	0.5.0-incubating	-	-
Hue	4.4.0	4.4.0	4.4.0	4.4.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.0.0	1.0.0	1.0.0	1.0.0
Livy	0.6.0	0.6.0	0.6.0	0.6.0
MXNet	1.5.1	1.5.1	1.4.0	1.4.0

	emr-5.28.1	emr-5.28.0	emr-5.27.1	emr-5.27.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.1.0	5.1.0	5.1.0	5.1.0
Phoenix	4.14.3	4.14.3	4.14.2	4.14.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.227	0.227	0.224	0.224
Spark	2.4.4	2.4.4	2.4.4	2.4.4
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.14.0	1.14.0	1.14.0	1.14.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.2	0.8.2	0.8.1	0.8.1
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

## 5.28.1 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.28.1. Änderungen beziehen sich auf Version 5.28.0.

Erste Version: 10. Januar 2020

Änderungen, Verbesserungen und behobene Probleme

- Spark
  - Spark-Kompatibilitätsprobleme behoben.
- CloudWatch Metriken

- Problem bei der Veröffentlichung von Amazon CloudWatch Metrics auf einem EMR-Cluster mit mehreren Primärknoten behoben.
- Deaktivierte Protokollnachricht
  - Falsche Protokollnachricht „... verwendet die alte Version (<4.5.8) des Apache-HTTP-Clients“ deaktiviert.

## Bekannte Probleme

- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in Amazon-EMR-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie dem Herunterskalieren oder der schrittweisen Übermittlung auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

### Workaround:

- SSH als hadoop-Benutzer für den führenden Primärknoten des EMR-Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den hadoop-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

### Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

## 5.28.1 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.2.6	Amazon SageMaker Spark-SDK
<code>emr-ddb</code>	4.12.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.11.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-s3-dist-cp</code>	2.13.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
<code>emr-s3-select</code>	1.3.0	EMR S3Select-Konnektor

Komponente	Version	Beschreibung
emrfs	2.37.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.9.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.5-amzn-5	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.5-amzn-5	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.5-amzn-5	HDFS-Client und -Bibliothek für die Befehlszeile

Komponente	Version	Beschreibung
hadoop-hdfs-namenode	2.8.5-amzn-5	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-hdfs-journalnode	2.8.5-amzn-5	HDFS-Service zum Verwalten des Hadoop-Dateisystem journals auf HA-Clustern.
hadoop-httfs-server	2.8.5-amzn-5	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.8.5-amzn-5	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.8.5-amzn-5	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.8.5-amzn-5	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.8.5-amzn-5	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.5-amzn-5	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.



Komponente	Version	Beschreibung
hbase-hmaster	1.4.10	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.4.10	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.4.10	HBase-Befehlszeilen-Client.
hbase-rest-server	1.4.10	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.4.10	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.6-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.6-amzn-0	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.3.6-amzn-0	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.3.6-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.3.6-amzn-0	Hive-hbase client.

Komponente	Version	Beschreibung
hive-metastore-server	2.3.6-amzn-0	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.3.6-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.5.0-incubating	Inkrementelles Verarbeitungs-Framework zur Stromversorgung der Datenpipeline bei geringer Latenz und hoher Effizienz.
hudi-presto	0.5.0-incubating	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hue-server	4.4.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	1.0.0	Multi-User-Server für Jupyter-Notebooks
livy-server	0.6.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP- und Reverse-Proxy-Server.
mahout-client	0.13.0	Bibliothek für Machine Learning.

Komponente	Version	Beschreibung
mxnet	1.5.1	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.1.0	Oozie-Befehlszeilen-Client.
oozie-server	5.1.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.3-HBase-1.4	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.14.3-HBase-1.4	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.227	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.227	Service für das Ausführen von Teilen einer Abfrage.

Komponente	Version	Beschreibung
presto-client	0.227	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.1	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
spark-client	2.4.4	Spark-Befehlszeilen-Clients.
spark-history-server	2.4.4	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.4.4	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.4.4	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	1.14.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.

Komponente	Version	Beschreibung
tez-on-yarn	0.9.2	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.8.2	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.14	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.14	ZooKeeper Befehlszeilen-Cli-ent.

## 5.28.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

### emr-5.28.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-log4j	Ändert die Werte in der <code>container-log4j.properties</code> -Datei in Hadoop YARN.

Klassifizierungen	Beschreibung
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4 j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.

Klassifizierungen	Beschreibung
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.

Klassifizierungen	Beschreibung
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.



Klassifizierungen	Beschreibung
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.

Klassifizierungen	Beschreibung
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.
ranger-kms-dbks-site	Ändert die Werte in der dbks-site.xml-Datei von Ranger KMS.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger KMS.
ranger-kms-env	Ändert die Werte in der Ranger KMS-Umgebung.

Klassifizierungen	Beschreibung
ranger-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei von Ranger KMS.
ranger-kms-db-ca	Ändert die Werte für die CA-Datei auf S3 für die MySQL SSL-Verbindung mit Ranger KMS.
recordserver-env	Werte in der RecordServer EMR-Umgebung ändern.
recordserver-conf	Ändern Sie die Werte in der Datei RecordServer server.properties von EMR.
recordserver-log4j	Ändern Sie die Werte in der Datei RecordServer log4j.properties von EMR.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml von Sqoop OraOop.

Klassifizierungen	Beschreibung
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.28.0

### 5.28.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)

- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

## Informationen zur Anwendungsversion

	emr-5.28.0	emr-5.27.1	emr-5.27.0	emr-5.26.0
AWS SDK for Java	1.11.659	1,11,615	1,11,615	1,11,595
Python	2,7, 3,6	2,7, 3,6	2,7, 3,6	2,7, 3,6
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.9.0	1.8.1	1.8.1	1.8.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.10	1.4.10	1.4.10	1.4.10
HCatalog	2.3.6	2.3.5	2.3.5	2.3.5
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.6	2.3.5	2.3.5	2.3.5
Hudi	0.5.0-incubating	-	-	-
Hue	4.4.0	4.4.0	4.4.0	4.4.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.0.0	1.0.0	1.0.0	0.9.6

	emr-5.28.0	emr-5.27.1	emr-5.27.0	emr-5.26.0
Livy	0.6.0	0.6.0	0.6.0	0.6.0
MXNet	1.5.1	1.4.0	1.4.0	1.4.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.1.0	5.1.0	5.1.0	5.1.0
Phoenix	4.14.3	4.14.2	4.14.2	4.14.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.227	0.224	0.224	0.220
Spark	2.4.4	2.4.4	2.4.4	2.4.3
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.14.0	1.14.0	1.14.0	1.13.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.2	0.8.1	0.8.1	0.8.1
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

## 5.28.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.28.0. Änderungen beziehen sich auf Version 5.27.0.

Erste Version: 12. November 2019

### Upgrades

- Upgrade von Flink auf Version 1.9.0

- Upgrade von Hive auf Version 2.3.6
- Upgrade von MXNet auf Version 1.5.1
- Upgrade von Phoenix auf Version 4.14.3
- Upgrade von Presto auf Version 0.227
- Upgrade von Zeppelin auf Version 0.8.2

## Neue Features

- [Apache Hudi](#) steht nun zur Installation unter Amazon EMR zur Verfügung, wenn Sie einen Cluster erstellen. Weitere Informationen finden Sie unter [Hudi](#).
- (25. November 2019) Sie können jetzt auf Wunsch mehrere Schritte gleichzeitig ausführen, um die Cluster-Nutzung zu verbessern und Kosten zu sparen. Außerdem können Sie sowohl anstehende als auch laufende Schritte stornieren. Weitere Informationen finden Sie unter [Arbeiten mit Schritten mithilfe der AWS CLI AND-Konsole](#).
- (3. Dezember 2019) Sie können jetzt EMR-Cluster erstellen und ausführen. AWS Outposts  
AWS Outposts ermöglicht native AWS Dienste, Infrastrukturen und Betriebsmodelle in lokalen Einrichtungen. In AWS Outposts Umgebungen können Sie dieselben AWS APIs, Tools und Infrastrukturen verwenden, die Sie in der AWS Cloud verwenden. Weitere Informationen finden Sie unter [EMR-Cluster auf AWS Outposts](#).
- (11. März 2020) Ab Amazon EMR Version 5.28.0 können Sie Amazon EMR-Cluster in einem Local Zones-Subnetz als logische Erweiterung einer AWS Region, die AWS Local Zones unterstützt, erstellen und ausführen. Eine lokale Zone ermöglicht es, Amazon EMR-Funktionen und eine Untergruppe von AWS Diensten, wie Rechen- und Speicherdienste, näher an den Benutzern zu platzieren, wodurch der Zugriff auf lokal ausgeführte Anwendungen mit sehr geringer Latenz ermöglicht wird. Eine Liste der verfügbaren Local Zones finden Sie unter [AWS Local Zones](#). Informationen zum Zugriff auf verfügbare AWS Local Zones finden Sie unter [Regionen, Availability Zones und Local Zones](#).

Local Zones unterstützen derzeit keine Amazon EMR Notebooks und unterstützen keine direkten Verbindungen zu Amazon EMR über die Schnittstelle VPC-Endpunkt (AWS PrivateLink).

## Änderungen, Verbesserungen und behobene Probleme

- Erweiterter Anwendungssupport für Cluster mit hoher Verfügbarkeit



- Weitere Informationen finden Sie unter [Unterstützte Anwendungen in einem EMR-Cluster mit mehreren Primärknoten](#) im Verwaltungshandbuch für Amazon EMR.
- Spark
  - Leistungsoptimierungen
- Hive
  - Leistungsoptimierungen
- Presto
  - Leistungsoptimierungen

## Bekannte Probleme

- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in Amazon-EMR-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie dem Herunterskalieren oder der schrittweisen Übermittlung auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

### Workaround:

- SSH als hadoop-Benutzer für den führenden Primärknoten des EMR-Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den hadoop-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

### Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen

konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

## 5.28.0 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.2.6	Amazon SageMaker Spark-SDK
<code>emr-ddb</code>	4.12.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.11.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-s3-dist-cp	2.13.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	1.3.0	EMR S3Select-Konnektor
emrfs	2.37.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.9.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.5-amzn-5	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.5-amzn-5	HDFS-Service auf Knotenebene zum Speichern von Blöcken.

Komponente	Version	Beschreibung
hadoop-hdfs-library	2.8.5-amzn-5	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.8.5-amzn-5	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-hdfs-journalnode	2.8.5-amzn-5	HDFS-Service zum Verwalten des Hadoop-Dateisystem journals auf HA-Clustern.
hadoop-httpfs-server	2.8.5-amzn-5	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.8.5-amzn-5	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.8.5-amzn-5	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.8.5-amzn-5	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.8.5-amzn-5	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.

Komponente	Version	Beschreibung
hadoop-yarn-timeline-server	2.8.5-amzn-5	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.4.10	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.4.10	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.4.10	HBase-Befehlszeilen-Client.
hbase-rest-server	1.4.10	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.4.10	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.6-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.6-amzn-0	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.3.6-amzn-0	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.

Komponente	Version	Beschreibung
hive-client	2.3.6-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.3.6-amzn-0	Hive-hbase client.
hive-metastore-server	2.3.6-amzn-0	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.3.6-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hudi	0.5.0-incubating	Inkrementelles Verarbeitungs-Framework zur Stromversorgung der Datenpipeline bei geringer Latenz und hoher Effizienz.
hudi-presto	0.5.0-incubating	Bundle-Bibliothek zum Ausführen von Presto mit Hudi.
hue-server	4.4.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	1.0.0	Multi-User-Server für Jupyter-Notebooks
livy-server	0.6.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP- und Reverse-Proxy-Server.

Komponente	Version	Beschreibung
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.5.1	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.1.0	Oozie-Befehlszeilen-Client.
oozie-server	5.1.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.3-HBase-1.4	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.14.3-HBase-1.4	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.227	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.227	Service für das Ausführen von Teilen einer Abfrage.

Komponente	Version	Beschreibung
presto-client	0.227	Presto-Befehlszeilenclient, der auf den Standby-Mastern eines HA-Clusters installiert ist, auf denen der Presto-Server nicht gestartet wird.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.1	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
spark-client	2.4.4	Spark-Befehlszeilen-Clients.
spark-history-server	2.4.4	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.4.4	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.4.4	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	1.14.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.



Komponente	Version	Beschreibung
tez-on-yarn	0.9.2	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.8.2	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.14	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.14	ZooKeeper Befehlszeilen-Cli-ent.

## 5.28.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

### emr-5.28.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-log4j	Ändert die Werte in der <code>container-log4j.properties</code> -Datei in Hadoop YARN.

Klassifizierungen	Beschreibung
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4 j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.

Klassifizierungen	Beschreibung
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.

Klassifizierungen	Beschreibung
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.

Klassifizierungen	Beschreibung
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.

Klassifizierungen	Beschreibung
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.
ranger-kms-dbks-site	Ändert die Werte in der dbks-site.xml-Datei von Ranger KMS.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger KMS.
ranger-kms-env	Ändert die Werte in der Ranger KMS-Umgebung.

Klassifizierungen	Beschreibung
ranger-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei von Ranger KMS.
ranger-kms-db-ca	Ändert die Werte für die CA-Datei auf S3 für die MySQL SSL-Verbindung mit Ranger KMS.
recordserver-env	Werte in der RecordServer EMR-Umgebung ändern.
recordserver-conf	Ändern Sie die Werte in der Datei RecordServer server.properties von EMR.
recordserver-log4j	Ändern Sie die Werte in der Datei RecordServer log4j.properties von EMR.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml von Sqoop OraOop.



Klassifizierungen	Beschreibung
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.27.1

### 5.27.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)

- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

## Informationen zur Anwendungsversion

	emr-5.27.1	emr-5.27.0	emr-5.26.0	emr-5.25.0
AWS SDK for Java	1.11.615	1,11,615	1,11,595	1,11,566
Python	2,7, 3,6	2,7, 3,6	2,7, 3,6	2,7, 3,6
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.8.1	1.8.1	1.8.0	1.8.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.10	1.4.10	1.4.10	1.4.9
HCatalog	2.3.5	2.3.5	2.3.5	2.3.5
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.5	2.3.5	2.3.5	2.3.5
Hudi	-	-	-	-
Hue	4.4.0	4.4.0	4.4.0	4.4.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.0.0	1.0.0	0.9.6	0.9.6

	emr-5.27.1	emr-5.27.0	emr-5.26.0	emr-5.25.0
Livy	0.6.0	0.6.0	0.6.0	0.6.0
MXNet	1.4.0	1.4.0	1.4.0	1.4.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.1.0	5.1.0	5.1.0	5.1.0
Phoenix	4.14.2	4.14.2	4.14.2	4.14.1
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.224	0.224	0.220	0.220
Spark	2.4.4	2.4.4	2.4.3	2.4.3
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.14.0	1.14.0	1.13.1	1.13.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.1	0.8.1	0.8.1	0.8.1
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

## 5.27.1 Versionshinweise

Dies ist eine Patch-Version. Alle Anwendungen und Komponenten sind dieselben wie in der vorherigen Amazon-EMR-Version.

Supportstatus für Instance Metadata Service (IMDS) V2: Amazon-EMR-Komponenten von 5.23.1, 5.27.1 und 5.32 oder höher verwenden IMDSv2 für alle IMDS-Aufrufe. Für IMDS-Aufrufe in Ihrem Anwendungscode können Sie sowohl IMDSv1 als auch IMDSv2 verwenden oder das IMDS so

konfigurieren, dass es aus Sicherheitsgründen nur IMDSv2 verwendet. Bei anderen 5.x-EMR-Versionen führt die Deaktivierung von IMDSv1 zu einem Cluster-Startup-Fehler.

## 5.27.1 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.2.4	Amazon SageMaker Spark-SDK
<code>emr-ddb</code>	4.12.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.11.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-s3-dist-cp</code>	2.13.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.

Komponente	Version	Beschreibung
emr-s3-select	1.3.0	EMR S3Select-Konnektor
emrfs	2.36.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.8.1	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.5-amzn-4	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.5-amzn-4	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.5-amzn-4	HDFS-Client und -Bibliothek für die Befehlszeile

Komponente	Version	Beschreibung
hadoop-hdfs-namenode	2.8.5-amzn-4	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-hdfs-journalnode	2.8.5-amzn-4	HDFS-Service zum Verwalten des Hadoop-Dateisystem journals auf HA-Clustern.
hadoop-httpfs-server	2.8.5-amzn-4	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.8.5-amzn-4	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.8.5-amzn-4	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.8.5-amzn-4	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resource-manager	2.8.5-amzn-4	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.5-amzn-4	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.

Komponente	Version	Beschreibung
hbase-hmaster	1.4.10	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.4.10	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.4.10	HBase-Befehlszeilen-Client.
hbase-rest-server	1.4.10	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.4.10	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.5-amzn-1	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.5-amzn-1	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.3.5-amzn-1	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.3.5-amzn-1	Hive-Befehlszeilen-Client.
hive-hbase	2.3.5-amzn-1	Hive-hbase client.

Komponente	Version	Beschreibung
hive-metastore-server	2.3.5-amzn-1	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.3.5-amzn-1	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.4.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	1.0.0	Multi-User-Server für Jupyter-Notebooks
livy-server	0.6.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP- und Reverse-Proxy-Server.
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.4.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Toolkit
oozie-client	5.1.0	Oozie-Befehlszeilen-Client.



Komponente	Version	Beschreibung
oozie-server	5.1.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.2-HBase-1.4	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.14.2-HBase-1.4	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.224	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.224	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.1	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
spark-client	2.4.4	Spark-Befehlszeilen-Clients.

Komponente	Version	Beschreibung
spark-history-server	2.4.4	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.4.4	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.4.4	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	1.14.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.2	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.8.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.14	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.

Komponente	Version	Beschreibung
zookeeper-client	3.4.14	ZooKeeper Befehlszeilen-Client.

### 5.27.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

#### emr-5.27.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-log4j	Ändert die Werte in der <code>container-log4j.properties</code> -Datei in Hadoop YARN.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Ändert die <code>log4j-cli.properties</code> -Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.

Klassifizierungen	Beschreibung
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site

Klassifizierungen	Beschreibung
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.

Klassifizierungen	Beschreibung
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.

Klassifizierungen	Beschreibung
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.

Klassifizierungen	Beschreibung
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.



Klassifizierungen	Beschreibung
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.
ranger-kms-dbks-site	Ändert die Werte in der dbks-site.xml-Datei von Ranger KMS.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger KMS.
ranger-kms-env	Ändert die Werte in der Ranger KMS-Umgebung.
ranger-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei von Ranger KMS.
ranger-kms-db-ca	Ändert die Werte für die CA-Datei auf S3 für die MySQL SSL-Verbindung mit Ranger KMS.
recordserver-env	Werte in der RecordServer EMR-Umgebung ändern.
recordserver-conf	Ändern Sie die Werte in der Datei RecordServer server.properties von EMR.
recordserver-log4j	Ändern Sie die Werte in der Datei RecordServer log4j.properties von EMR.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.

Klassifizierungen	Beschreibung
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml von Sqoop OraOop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.27.0

### 5.27.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

#### Informationen zur Anwendungsversion

	emr-5.27.0	emr-5.26.0	emr-5.25.0	emr-5.24.1
AWS SDK for Java	1.11.615	1,11.595	1,11.566	1,11.546
Python	2,7, 3,6	2,7, 3,6	2,7, 3,6	2,7, 3,6
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.8.1	1.8.0	1.8.0	1.8.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2

	emr-5.27.0	emr-5.26.0	emr-5.25.0	emr-5.24.1
HBase	1.4.10	1.4.10	1.4.9	1.4.9
HCatalog	2.3.5	2.3.5	2.3.5	2.3.4
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.5	2.3.5	2.3.5	2.3.4
Hudi	-	-	-	-
Hue	4.4.0	4.4.0	4.4.0	4.4.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.0.0	0.9.6	0.9.6	0.9.6
Livy	0.6.0	0.6.0	0.6.0	0.6.0
MXNet	1.4.0	1.4.0	1.4.0	1.4.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.1.0	5.1.0	5.1.0	5.1.0
Phoenix	4.14.2	4.14.2	4.14.1	4.14.1
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.224	0.220	0.220	0.219
Spark	2.4.4	2.4.3	2.4.3	2.4.2
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.14.0	1.13.1	1.13.1	1.12.0
Tez	0.9.2	0.9.2	0.9.2	0.9.1

	emr-5.27.0	emr-5.26.0	emr-5.25.0	emr-5.24.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.1	0.8.1	0.8.1	0.8.1
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.13

## 5.27.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.27.0. Änderungen beziehen sich auf Version 5.26.0.

Erste Version: 23. September 2019

### Upgrades

- AWS SDK for Java 1,11.615
- Flink 1.8.1
- JupyterHub 1.0.0
- Spark 2.4.4
- Tensorflow 1,14.0
- Konnektoren und Treiber:
  - DynamoDB Connector 4.12.0

### Neue Features

- (24. Oktober 2019) Die folgenden neuen Features in EMR Notebooks sind in allen Amazon-EMR-Versionen verfügbar.
  - Sie können Git-Repositorys mit Ihren EMR-Notebooks verknüpfen, um Ihre Notebooks in einer versionskontrollierten Umgebung zu speichern. Sie können Code mit Kollegen teilen und vorhandene Jupyter Notebooks über Remote-Git-Repositorys wiederverwenden. Weitere Informationen finden Sie unter [Verknüpfen von Git-basierten Repositorys mit Amazon EMR Notebooks](#) im Verwaltungshandbuch für Amazon EMR.

- Das [Hilfsprogramm nbdime](#) ist jetzt in EMR Notebooks verfügbar, um das Vergleichen und Zusammenführen von Notebooks zu vereinfachen.
- EMR-Notebooks werden jetzt unterstützt JupyterLab. JupyterLab ist eine webbasierte interaktive Entwicklungsumgebung, die vollständig mit Jupyter-Notebooks kompatibel ist. Sie können jetzt wählen, ob Sie Ihr Notizbuch entweder im JupyterLab Jupyter-Notebook-Editor öffnen möchten.
- (30. Oktober 2019) Mit Amazon-EMR-Version 5.25.0 können Sie über die Seite Zusammenfassung des Clusters oder die Registerkarte Anwendungsverlauf in der Konsole eine Verbindung zu persistenten Spark History Servern herstellen. Anstatt einen Web-Proxy über eine SSH-Verbindung einzurichten, können Sie schnell auf die Benutzeroberfläche des Spark-History-Servers zugreifen, um Anwendungsmetriken einzusehen und auf relevante Protokolldateien für aktive und beendete Cluster zuzugreifen. Weitere Informationen finden Sie unter [Off-Cluster-Zugriff auf persistente Anwendungsbenuzeroberflächen](#) im Verwaltungshandbuch für Amazon EMR.

## Änderungen, Verbesserungen und behobene Probleme

- Amazon-EMR-Cluster mit mehreren Primärknoten
  - Sie können Flink auf einem Amazon-EMR-Cluster mit mehreren Primärknoten installieren und ausführen. Weitere Informationen finden Sie unter [Unterstützte Anwendungen und Features](#).
  - Sie können die transparente HDFS-Verschlüsselung auf einem Amazon-EMR-Cluster mit mehreren Primärknoten konfigurieren. Weitere Informationen finden Sie unter [Transparente HDFS-Verschlüsselung auf EMR-Clustern mit mehreren Primärknoten](#).
  - Sie können jetzt die Konfiguration von Anwendungen ändern, die auf einem Amazon-EMR-Cluster mit mehreren Primärknoten ausgeführt werden. Weitere Informationen finden Sie unter [Angaben einer Konfiguration für eine Instance-Gruppe in einem aktiven Cluster](#).
- Amazon-EMR-DynamoDB-Konnektor
  - Amazon-EMR-DynamoDB-Konnektor unterstützt jetzt die folgenden DynamoDB-Datentypen: boolean, list, map, item, null. Weitere Informationen finden Sie unter [Eine Hive-Tabelle zur Ausführung von Hive-Befehlen einrichten](#).

## Bekannte Probleme

- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in Amazon-EMR-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie dem

Herunterskalieren oder der schrittweisen Übermittlung auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

Workaround:

- SSH als hadoop-Benutzer für den führenden Primärknoten des EMR-Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den hadoop-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

#### Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

## 5.27.0 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-

Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.2.4	Amazon SageMaker Spark-SDK
<code>emr-ddb</code>	4.12.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.11.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-s3-dist-cp</code>	2.13.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
<code>emr-s3-select</code>	1.3.0	EMR S3Select-Konnektor
<code>emrfs</code>	2.36.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>flink-client</code>	1.8.1	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
<code>ganglia-monitor</code>	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.



Komponente	Version	Beschreibung
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.5-amzn-4	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.5-amzn-4	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.5-amzn-4	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.8.5-amzn-4	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-hdfs-journalnode	2.8.5-amzn-4	HDFS-Service zum Verwalten des Hadoop-Dateisystem journals auf HA-Clustern.
hadoop-httpfs-server	2.8.5-amzn-4	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.8.5-amzn-4	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider

Komponente	Version	Beschreibung
hadoop-mapred	2.8.5-amzn-4	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.8.5-amzn-4	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.8.5-amzn-4	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.5-amzn-4	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.4.10	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.4.10	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.4.10	HBase-Befehlszeilen-Client.
hbase-rest-server	1.4.10	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.4.10	Service, der einen Thrift-Endpunkt für HBase bereitstellt.

Komponente	Version	Beschreibung
hcatalog-client	2.3.5-amzn-1	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.5-amzn-1	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.3.5-amzn-1	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.3.5-amzn-1	Hive-Befehlszeilen-Client.
hive-hbase	2.3.5-amzn-1	Hive-hbase client.
hive-metastore-server	2.3.5-amzn-1	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.3.5-amzn-1	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.4.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	1.0.0	Multi-User-Server für Jupyter-Notebooks
livy-server	0.6.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark

Komponente	Version	Beschreibung
nginx	1.12.1	nginx [engine x] ist ein HTTP- und Reverse-Proxy-Server.
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.4.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.1.0	Oozie-Befehlszeilen-Client.
oozie-server	5.1.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.2-HBase-1.4	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.14.2-HBase-1.4	Ein schlanker Server für den Zugriff auf JDBC und Protokoll puffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.224	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.

Komponente	Version	Beschreibung
presto-worker	0.224	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.1	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
spark-client	2.4.4	Spark-Befehlszeilen-Clients.
spark-history-server	2.4.4	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.4.4	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.4.4	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	1.14.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.2	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.

Komponente	Version	Beschreibung
zeppelin-server	0.8.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.14	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.14	ZooKeeper Befehlszeilen-Cli-ent.

## 5.27.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

### emr-5.27.0 Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-log4j	Ändert die Werte in der <code>container-log4j.properties</code> -Datei in Hadoop YARN.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.

Klassifizierungen	Beschreibung
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4 j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.

Klassifizierungen	Beschreibung
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.



Klassifizierungen	Beschreibung
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.

Klassifizierungen	Beschreibung
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.

Klassifizierungen	Beschreibung
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.
ranger-kms-dbks-site	Ändert die Werte in der dbks-site.xml-Datei von Ranger KMS.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger KMS.
ranger-kms-env	Ändert die Werte in der Ranger KMS-Umgebung.
ranger-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei von Ranger KMS.
ranger-kms-db-ca	Ändert die Werte für die CA-Datei auf S3 für die MySQL SSL-Verbindung mit Ranger KMS.
recordserver-env	Werte in der RecordServer EMR-Umgebung ändern.

Klassifizierungen	Beschreibung
recordserver-conf	Ändern Sie die Werte in der Datei RecordServer server.properties von EMR.
recordserver-log4j	Ändern Sie die Werte in der Datei RecordServer log4j.properties von EMR.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml von Sqoop OraOop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.

Klassifizierungen	Beschreibung
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.26.0

### 5.26.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

## Informationen zur Anwendungsversion

	emr-5.26.0	emr-5.25.0	emr-5.24.1	emr-5.24.0
AWS SDK for Java	1.11.595	1,11.566	1,11.546	1,11.546
Python	2,7, 3,6	2,7, 3,6	2,7, 3,6	2,7, 3,6
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.8.0	1.8.0	1.8.0	1.8.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.10	1.4.9	1.4.9	1.4.9
HCatalog	2.3.5	2.3.5	2.3.4	2.3.4
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.5	2.3.5	2.3.4	2.3.4
Hudi	-	-	-	-
Hue	4.4.0	4.4.0	4.4.0	4.4.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.9.6	0.9.6	0.9.6	0.9.6
Livy	0.6.0	0.6.0	0.6.0	0.6.0
MXNet	1.4.0	1.4.0	1.4.0	1.4.0

	emr-5.26.0	emr-5.25.0	emr-5.24.1	emr-5.24.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.1.0	5.1.0	5.1.0	5.1.0
Phoenix	4.14.2	4.14.1	4.14.1	4.14.1
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.220	0.220	0.219	0.219
Spark	2.4.3	2.4.3	2.4.2	2.4.2
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.13.1	1.13.1	1.12.0	1.12.0
Tez	0.9.2	0.9.2	0.9.1	0.9.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.1	0.8.1	0.8.1	0.8.1
ZooKeeper	3.4.14	3.4.14	3.4.13	3.4.13

## 5.26.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.26.0. Änderungen beziehen sich auf Version 5.25.0.

Erste Version: 8. August 2019

Letzte Aktualisierung: 19. August 2019

### Upgrades

- AWS SDK for Java 1,11.595
- HBase 1.4.10
- Phoenix 4.14.2



- Konnektoren und Treiber:
  - DynamoDB Connector 4.11.0
  - MariaDB Connector 2.4.2
  - Amazon-Redshift-JDBC-Treiber 1.2.32.1056

## Neue Features

- (Beta) Mit Amazon EMR 5.26.0 können Sie einen Cluster starten, der in Lake Formation integriert ist. Diese Integration bietet einen detaillierten Zugriff auf Spaltenebene auf Datenbanken und Tabellen im Glue-Datenkatalog. AWS Es ermöglicht Lake Formation über ein Unternehmens-Identitätssystem eine verbundene einmalige Anmeldung bei EMR Notebooks oder Apache Zeppelin. Weitere Informationen finden Sie unter [Integration von Amazon EMR mit AWS Lake Formation \(Beta\)](#).
- (19. August 2019) Amazon EMR Block Public Access ist jetzt in allen Amazon-EMR-Versionen verfügbar, die Sicherheitsgruppen unterstützen. Den öffentlichen Zugriff blockieren ist eine kontoweite Einstellung, die für jede Region gilt. AWS Block Public Access verhindert das Starten eines Clusters, wenn eine mit dem Cluster verknüpfte Sicherheitsgruppe über eine Regel verfügt, die eingehenden Datenverkehr von IPv4 0.0.0.0/0 oder IPv6 ::/0 (öffentlicher Zugriff) auf einem Port zulässt, sofern der Port nicht als Ausnahme festgelegt wurde. Port 22 ist standardmäßig eine Ausnahme. Weitere Informationen finden Sie unter [Verwenden von Amazon EMR Block Public Access](#) im Verwaltungshandbuch für Amazon EMR.

## Änderungen, Verbesserungen und behobene Probleme

- EMR Notebooks
  - Mit EMR 5.26.0 und höher unterstützt EMR Notebooks zusätzlich zu den Standard-Python-Bibliotheken auch Python-Bibliotheken für Notebooks. Sie können Bibliotheken für Notebooks vom Notebook-Editor aus installieren, ohne einen Cluster neu erstellen oder ein Notebook erneut an einen Cluster anhängen zu müssen. Notebook-bezogene Bibliotheken werden in einer virtuellen Python-Umgebung erstellt und gelten daher nur für die aktuelle Notebook-Sitzung. Auf diese Weise können Sie Abhängigkeiten zwischen Notebooks isolieren. Weitere Informationen finden Sie unter [Verwenden von Notebook Scoped Libraries](#) im Verwaltungshandbuch für Amazon EMR.
- EMRFS

- Sie können ein ETag-Überprüfungsfeature (Beta) aktivieren, indem Sie `fs.s3.consistent.metadata.etag verification.enabled` auf `true` einstellen. Mit diesem Feature verwendet EMRFS Amazon-S3-ETags, um zu überprüfen, ob gelesene Objekte die neueste verfügbare Version sind. Diese Funktion ist hilfreich für `read-after-update` Anwendungsfälle, in denen Dateien auf Amazon S3 überschrieben werden, während derselbe Name beibehalten wird. Diese ETag-Verifizierungsfunktion funktioniert derzeit nicht mit S3 Select. Weitere Informationen finden Sie unter [Konfigurieren der konsistenten Ansicht](#).
- Spark
  - Die folgenden Optimierungen sind jetzt standardmäßig aktiviert: dynamisches Bereinigen von Partitionen, DISTINCT vor INTERSECT, Verbesserungen bei der Inferenz von SQL-Planstatistiken für JOIN, gefolgt von DISTINCT-Abfragen, Reduzierung skalarer Unterabfragen, optimierte Join-Neuordnung und Bloom-Filter-Join. Weitere Informationen finden Sie unter [Optimieren der Spark-Leistung](#).
  - Die Codegenerierung für die gesamte Phase für Sort Merge Join wurde verbessert.
  - Die Wiederverwendung von Abfragefragmenten und Unterabfragen wurde verbessert.
  - Verbesserungen bei der Vorabzuweisung von Executoren beim Start von Spark.
  - Bloom-Filter-Joins werden nicht mehr angewendet, wenn die kleinere Seite der Verknüpfung einen Broadcast-Hinweis enthält.
- Tez
  - Ein Problem mit Tez wurde behoben. Die Tez-Benutzeroberfläche funktioniert jetzt auf einem Amazon-EMR-Cluster mit mehreren Primärknoten.

## Bekannte Probleme

- Die verbesserten Funktionen zur Codegenerierung in der gesamten Phase für Sort Merge Join können den Speicherdruck erhöhen, wenn sie aktiviert sind. Diese Optimierung verbessert die Leistung, kann jedoch zu Wiederholungsversuchen oder Fehlschlägen führen, wenn der nicht so eingestellt `spark.yarn.executor.memoryOverheadFactor` ist, dass er genügend Speicher zur Verfügung stellt. Um dieses Feature zu deaktivieren, setzen Sie `spark.sql.sortMergeJoinExec.extendedCodegen.enabled` auf „false“.
- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in Amazon-EMR-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie dem Herunterskalieren oder der schrittweisen Übermittlung auftreten, nachdem der Cluster einige Zeit

lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

Workaround:

- SSH als hadoop-Benutzer für den führenden Primärknoten des EMR-Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den hadoop-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

#### Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

## 5.26.0 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die

Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als 2.2-amzn-2 aufgeführt.

Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1.2.4	Amazon SageMaker Spark-SDK
emr-ddb	4.11.0	Amazon DynamoDB-Konnektor für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.10.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.12.0	Für Amazon S3 optimierte verteilte Kopieranwendung.
emr-s3-select	1.3.0	EMR S3Select-Konnektor
emrfs	2.35.0	Amazon S3-Konnektor für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.8.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.

Komponente	Version	Beschreibung
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.5-amzn-4	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.5-amzn-4	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.5-amzn-4	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.8.5-amzn-4	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-hdfs-journalnode	2.8.5-amzn-4	HDFS-Service zum Verwalten des Hadoop-Dateisystem journals auf HA-Clustern.
hadoop-httpfs-server	2.8.5-amzn-4	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.8.5-amzn-4	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider

Komponente	Version	Beschreibung
hadoop-mapred	2.8.5-amzn-4	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.8.5-amzn-4	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.8.5-amzn-4	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.5-amzn-4	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.4.10	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.4.10	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.4.10	HBase-Befehlszeilen-Client.
hbase-rest-server	1.4.10	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.4.10	Service, der einen Thrift-Endpunkt für HBase bereitstellt.

Komponente	Version	Beschreibung
hcatalog-client	2.3.5-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.5-amzn-0	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.3.5-amzn-0	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.3.5-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.3.5-amzn-0	Hive-hbase client.
hive-metastore-server	2.3.5-amzn-0	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.3.5-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.4.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	0.9.6	Multi-User-Server für Jupyter-Notebooks
livy-server	0.6.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark

Komponente	Version	Beschreibung
nginx	1.12.1	nginx [engine x] ist ein HTTP- und Reverse-Proxy-Server.
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.4.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.1.0	Oozie-Befehlszeilen-Client.
oozie-server	5.1.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.2-HBase-1.4	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.14.2-HBase-1.4	Ein schlanker Server für den Zugriff auf JDBC und Protokoll puffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.220	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.



Komponente	Version	Beschreibung
presto-worker	0.220	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.1	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
spark-client	2.4.3	Spark-Befehlszeilen-Clients.
spark-history-server	2.4.3	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.4.3	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.4.3	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	1.13.1	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.2	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.

Komponente	Version	Beschreibung
zeppelin-server	0.8.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.14	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.14	ZooKeeper Befehlszeilen-Cli-ent.

## 5.26.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

### emr-5.26.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-log4j	Ändert die Werte in der <code>container-log4j.properties</code> -Datei in Hadoop YARN.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.

Klassifizierungen	Beschreibung
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4 j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.

Klassifizierungen	Beschreibung
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.

Klassifizierungen	Beschreibung
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.

Klassifizierungen	Beschreibung
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.

Klassifizierungen	Beschreibung
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.
recordserver-env	Werte in der RecordServer EMR-Umgebung ändern.
recordserver-conf	Ändern Sie die Werte in der Datei RecordServer server.properties von EMR.
recordserver-log4j	Ändern Sie die Werte in der Datei RecordServer log4j.properties von EMR.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.



Klassifizierungen	Beschreibung
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml von Sqoop OraOop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.25.0

### 5.25.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

#### Informationen zur Anwendungsversion

	emr-5.25.0	emr-5.24.1	emr-5.24.0	emr-5.23.1
AWS SDK for Java	1.11.566	1,11.546	1,11.546	1,11.519
Python	2,7, 3,6	2,7, 3,6	2,7, 3,6	2,7, 3,6
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.8.0	1.8.0	1.8.0	1.7.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2

	emr-5.25.0	emr-5.24.1	emr-5.24.0	emr-5.23.1
HBase	1.4.9	1.4.9	1.4.9	1.4.9
HCatalog	2.3.5	2.3.4	2.3.4	2.3.4
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.5	2.3.4	2.3.4	2.3.4
Hudi	-	-	-	-
Hue	4.4.0	4.4.0	4.4.0	4.3.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.9.6	0.9.6	0.9.6	0.9.4
Livy	0.6.0	0.6.0	0.6.0	0.5.0
MXNet	1.4.0	1.4.0	1.4.0	1.3.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.1.0	5.1.0	5.1.0	5.1.0
Phoenix	4.14.1	4.14.1	4.14.1	4.14.1
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.220	0.219	0.219	0.215
Spark	2.4.3	2.4.2	2.4.2	2.4.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.13.1	1.12.0	1.12.0	1.12.0
Tez	0.9.2	0.9.1	0.9.1	0.9.1

	emr-5.25.0	emr-5.24.1	emr-5.24.0	emr-5.23.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.1	0.8.1	0.8.1	0.8.1
ZooKeeper	3.4.14	3.4.13	3.4.13	3.4.13

## 5.25.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.25.0. Änderungen beziehen sich auf Version 5.24.1.

Erste Version: 17. Juli 2019

Letzte Aktualisierung: 30. Oktober 2019

### Amazon EMR 5.25.0

#### Upgrades

- AWS SDK for Java 1,11.566
- Hive 2.3.5
- Presto 0.220
- Spark 2.4.3
- TensorFlow 1,13,1
- Tez 0.9.2
- Zookeeper 3.4.14

#### Neue Features

- (30. Oktober 2019) Beginnend mit der Amazon-EMR-Version 5.25.0 können Sie über die Seite Zusammenfassung des Clusters oder die Registerkarte Anwendungsverlauf in der Konsole eine Verbindung zu persistenten Spark-History-Servern herstellen. Anstatt einen Web-Proxy über eine SSH-Verbindung einzurichten, können Sie schnell auf die Benutzeroberfläche des Spark-History-Servers zugreifen, um Anwendungsmetriken einzusehen und auf relevante Protokolldateien für aktive und beendete Cluster zuzugreifen. Weitere Informationen finden Sie unter [Off-Cluster-](#)

[Zugriff auf persistente Anwendungsbenutzeroberflächen](#) im Verwaltungshandbuch für Amazon EMR.

## Änderungen, Verbesserungen und behobene Probleme

- Spark
  - Die Leistung einiger Verknüpfungen wurde verbessert, indem Bloom-Filter zum Vorfiltern von Eingaben verwendet wurden. Die Optimierung ist standardmäßig deaktiviert und kann aktiviert werden, indem der Spark-Konfigurationsparameter `spark.sql.bloomFilterJoin.enabled` auf `true` gesetzt wird.
  - Die Leistung bei der Gruppierung nach Spalten vom Typ Zeichenfolge wurde verbessert.
  - Die standardmäßige Speicher- und Kernkonfiguration des Spark-Executors von R4-Instance-Typen für Cluster ohne installiertes HBase wurde verbessert.
  - Es wurde ein früheres Problem mit dem Feature zum dynamischen Bereinigen von Partitionen behoben, bei dem sich die bereinigte Tabelle auf der linken Seite des Joins befinden musste.
  - Die Optimierung von DISTINCT vor INTERSECT wurde verbessert und gilt nun auch für weitere Fälle, in denen Aliase involviert sind.
  - Die Inferenz der SQL-Planstatistiken für JOIN, gefolgt von DISTINCT-Abfragen, wurde verbessert. Diese Verbesserung ist standardmäßig deaktiviert und kann aktiviert werden, indem der Spark-Konfigurationsparameter `spark.sql.statsImprovements.enabled` auf `true` gesetzt wird. Diese Optimierung ist für das Feature „Unterscheiden vor Überschneiden“ erforderlich und wird automatisch aktiviert, wenn `spark.sql.optimizer.distinctBeforeIntersect.enabled` auf `true` gesetzt ist.
  - Optimierte Verbindungsreihenfolge basierend auf Tabellengröße und Filtern. Diese Optimierung ist standardmäßig deaktiviert und kann aktiviert werden, indem der Spark-Konfigurationsparameter `spark.sql.optimizer.sizeBasedJoinReorder.enabled` auf `true` gesetzt wird.

Weitere Informationen finden Sie unter [Optimieren der Spark-Leistung](#).

- EMRFS
  - Die EMRFS-Einstellung `fs.s3.buckets.create.enabled` ist jetzt standardmäßig deaktiviert. Bei Tests haben wir festgestellt, dass die Deaktivierung dieser Einstellung die Leistung verbessert und die unbeabsichtigte Erstellung von S3-Buckets verhindert. Wenn Ihre Anwendung auf dieser Funktionalität basiert, können Sie sie aktivieren, indem Sie die Eigenschaft `true` in der Konfigurationsklassifizierung von `fs.s3.buckets.create.enabled`

auf `emrfs-site` setzen. Weitere Informationen finden Sie unter [Angeben einer Konfiguration beim Erstellen eines Clusters](#).

- Verbesserungen der lokalen Festplattenverschlüsselung und der S3-Verschlüsselung in Sicherheitskonfigurationen (5. August 2019)
  - Die Amazon-S3-Verschlüsselungseinstellungen wurden im Sicherheitskonfigurations-Setup von den Einstellungen für die lokale Festplattenverschlüsselung getrennt.
  - Es wurde eine Option hinzugefügt, um die EBS-Verschlüsselung mit Version 5.24.0 und höher zu aktivieren. Bei Auswahl dieser Option wird zusätzlich zu den Speichervolumen auch das Root-Geräte-Volumen verschlüsselt. Frühere Versionen erforderten die Verwendung eines benutzerdefinierten AMI, um das Root-Geräte-Volumen zu verschlüsseln.
  - Weitere Informationen erhalten Sie unter [Verschlüsselungsoptionen](#) im Verwaltungshandbuch für Amazon EMR.

## Bekannte Probleme

- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in Amazon-EMR-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie dem Herunterskalieren oder der schrittweisen Übermittlung auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

### Workaround:

- SSH als `hadoop`-Benutzer für den führenden Primärknoten des EMR-Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den `hadoop`-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

### Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

## 5.25.0 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.2.4	Amazon SageMaker Spark-SDK
<code>emr-ddb</code>	4.10.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.9.0	Praktische Bibliotheken für das Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-kinesis	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.11.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	1.3.0	EMR S3Select-Konnektor
emrfs	2.34.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.8.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.5-amzn-4	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".



Komponente	Version	Beschreibung
hadoop-hdfs-datanode	2.8.5-amzn-4	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.5-amzn-4	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.8.5-amzn-4	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-hdfs-journalnode	2.8.5-amzn-4	HDFS-Service zum Verwalten des Hadoop-Dateisystem journals auf HA-Clustern.
hadoop-https-server	2.8.5-amzn-4	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.8.5-amzn-4	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.8.5-amzn-4	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.8.5-amzn-4	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.8.5-amzn-4	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.

Komponente	Version	Beschreibung
hadoop-yarn-timeline-server	2.8.5-amzn-4	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.4.9	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.4.9	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.4.9	HBase-Befehlszeilen-Client.
hbase-rest-server	1.4.9	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.4.9	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.5-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.5-amzn-0	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.3.5-amzn-0	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.

Komponente	Version	Beschreibung
hive-client	2.3.5-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.3.5-amzn-0	Hive-hbase client.
hive-metastore-server	2.3.5-amzn-0	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.3.5-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.4.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	0.9.6	Multi-User-Server für Jupyter-Notebooks
livy-server	0.6.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP- und Reverse-Proxy-Server.
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.4.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.

Komponente	Version	Beschreibung
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.1.0	Oozie-Befehlszeilen-Client.
oozie-server	5.1.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.1-HBase-1.4	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.14.1-HBase-1.4	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.220	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.220	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.1	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
spark-client	2.4.3	Spark-Befehlszeilen-Clients.

Komponente	Version	Beschreibung
spark-history-server	2.4.3	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.4.3	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.4.3	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	1.13.1	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.2	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.8.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.14	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.

Komponente	Version	Beschreibung
zookeeper-client	3.4.14	ZooKeeper Befehlszeilen-Client.

## 5.25.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

### emr-5.25.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-log4j	Ändert die Werte in der <code>container-log4j.properties</code> -Datei in Hadoop YARN.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Ändert die <code>log4j-cli.properties</code> -Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.

Klassifizierungen	Beschreibung
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site

Klassifizierungen	Beschreibung
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.



Klassifizierungen	Beschreibung
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.

Klassifizierungen	Beschreibung
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.

Klassifizierungen	Beschreibung
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.
recordserver-env	Werte in der RecordServer EMR-Umgebung ändern.
recordserver-conf	Ändern Sie die Werte in der Datei RecordServer server.properties von EMR.
recordserver-log4j	Ändern Sie die Werte in der Datei RecordServer log4j.properties von EMR.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.

Klassifizierungen	Beschreibung
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml von Sqoop OraOop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.24.1

### 5.24.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)

- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

### Informationen zur Anwendungsversion

	emr-5.24.1	emr-5.24.0	emr-5.23.1	emr-5.23.0
AWS SDK for Java	1.11.546	1,11.546	1,11.519	1,11.519
Python	2,7, 3,6	2,7, 3,6	2,7, 3,6	2,7, 3,6
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.8.0	1.8.0	1.7.1	1.7.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.9	1.4.9	1.4.9	1.4.9
HCatalog	2.3.4	2.3.4	2.3.4	2.3.4
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.4	2.3.4	2.3.4	2.3.4
Hudi	-	-	-	-
Hue	4.4.0	4.4.0	4.3.0	4.3.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-

	emr-5.24.1	emr-5.24.0	emr-5.23.1	emr-5.23.0
JupyterHub	0.9.6	0.9.6	0.9.4	0.9.4
Livy	0.6.0	0.6.0	0.5.0	0.5.0
MXNet	1.4.0	1.4.0	1.3.1	1.3.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.1.0	5.1.0	5.1.0	5.1.0
Phoenix	4.14.1	4.14.1	4.14.1	4.14.1
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.219	0.219	0.215	0.215
Spark	2.4.2	2.4.2	2.4.0	2.4.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.12.0	1.12.0	1.12.0	1.12.0
Tez	0.9.1	0.9.1	0.9.1	0.9.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.1	0.8.1	0.8.1	0.8.1
ZooKeeper	3.4.13	3.4.13	3.4.13	3.4.13

## 5.24.1 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.24.1. Änderungen beziehen sich auf Version 5.24.0.

Erste Version: 26. Juni 2019

## Änderungen, Verbesserungen und behobene Probleme

- Das standardmäßige Amazon Linux AMI für Amazon EMR wurde aktualisiert, sodass es wichtige Sicherheitsupdates für den Linux-Kernel enthält, darunter das „TCP SACK Denial of Service Issue“ ([AWS-2019-005](#)).

### Bekannte Probleme

- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in Amazon-EMR-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie dem Herunterskalieren oder der schrittweisen Übermittlung auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

#### Workaround:

- SSH als hadoop-Benutzer für den führenden Primärknoten des EMR-Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den hadoop-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

#### Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.



## 5.24.1 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.2.1	Amazon SageMaker Spark-SDK
<code>emr-ddb</code>	4.9.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.8.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-s3-dist-cp</code>	2.11.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
<code>emr-s3-select</code>	1.3.0	EMR S3Select-Konnektor

Komponente	Version	Beschreibung
emrfs	2.33.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.8.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.5-amzn-4	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.5-amzn-4	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.5-amzn-4	HDFS-Client und -Bibliothek für die Befehlszeile

Komponente	Version	Beschreibung
hadoop-hdfs-namenode	2.8.5-amzn-4	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-hdfs-journalnode	2.8.5-amzn-4	HDFS-Service zum Verwalten des Hadoop-Dateisystem journals auf HA-Clustern.
hadoop-httfs-server	2.8.5-amzn-4	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.8.5-amzn-4	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.8.5-amzn-4	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.8.5-amzn-4	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resource-manager	2.8.5-amzn-4	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.5-amzn-4	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.

Komponente	Version	Beschreibung
hbase-hmaster	1.4.9	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.4.9	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.4.9	HBase-Befehlszeilen-Client.
hbase-rest-server	1.4.9	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.4.9	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.4-amzn-2	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.4-amzn-2	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.3.4-amzn-2	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.3.4-amzn-2	Hive-Befehlszeilen-Client.
hive-hbase	2.3.4-amzn-2	Hive-hbase client.

Komponente	Version	Beschreibung
hive-metastore-server	2.3.4-amzn-2	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.3.4-amzn-2	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.4.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	0.9.6	Multi-User-Server für Jupyter-Notebooks
livy-server	0.6.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP- und Reverse-Proxy-Server.
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.4.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Toolkit
oozie-client	5.1.0	Oozie-Befehlszeilen-Client.

Komponente	Version	Beschreibung
oozie-server	5.1.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.1-HBase-1.4	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.14.1-HBase-1.4	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.219	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.219	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.1	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
spark-client	2.4.2	Spark-Befehlszeilen-Clients.

Komponente	Version	Beschreibung
spark-history-server	2.4.2	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.4.2	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.4.2	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	1.12.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.1	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.8.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.13	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.

Komponente	Version	Beschreibung
zookeeper-client	3.4.13	ZooKeeper Befehlszeilen-Client.

## 5.24.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

### emr-5.24.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-log4j	Ändert die Werte in der <code>container-log4j.properties</code> -Datei in Hadoop YARN.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Ändert die <code>log4j-cli.properties</code> -Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.



Klassifizierungen	Beschreibung
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site

Klassifizierungen	Beschreibung
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.

Klassifizierungen	Beschreibung
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.

Klassifizierungen	Beschreibung
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.

Klassifizierungen	Beschreibung
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.

Klassifizierungen	Beschreibung
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.24.0

### 5.24.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

## Informationen zur Anwendungsversion

	emr-5.24.0	emr-5.23.1	emr-5.23.0	emr-5.22.0
AWS SDK for Java	1.11.546	1,11.519	1,11.519	1,11.510
Python	2,7, 3,6	2,7, 3,6	2,7, 3,6	2,7, 3,6
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.8.0	1.7.1	1.7.1	1.7.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.9	1.4.9	1.4.9	1.4.9
HCatalog	2.3.4	2.3.4	2.3.4	2.3.4
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.4	2.3.4	2.3.4	2.3.4
Hudi	-	-	-	-
Hue	4.4.0	4.3.0	4.3.0	4.3.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.9.6	0.9.4	0.9.4	0.9.4
Livy	0.6.0	0.5.0	0.5.0	0.5.0
MXNet	1.4.0	1.3.1	1.3.1	1.3.1



	emr-5.24.0	emr-5.23.1	emr-5.23.0	emr-5.22.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.1.0	5.1.0	5.1.0	5.1.0
Phoenix	4.14.1	4.14.1	4.14.1	4.14.1
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.219	0.215	0.215	0.215
Spark	2.4.2	2.4.0	2.4.0	2.4.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.12.0	1.12.0	1.12.0	1.12.0
Tez	0.9.1	0.9.1	0.9.1	0.9.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.1	0.8.1	0.8.1	0.8.1
ZooKeeper	3.4.13	3.4.13	3.4.13	3.4.13

## 5.24.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.24.0. Änderungen beziehen sich auf Version 5.23.0.

Erste Version: 11. Juni 2019

Letzte Aktualisierung: 5. August 2019

### Upgrades

- Flink 1.8.0
- Hue 4.4.0
- JupyterHub 0.9.6

- Livy 0.6.0
- MxNet 1.4.0
- Presto 0.219
- Spark 2.4.2
- AWS SDK for Java 1,11.546
- Konnektoren und Treiber:
  - DynamoDB Connector 4.9.0
  - MariaDB Connector 2.4.1
  - Amazon-Redshift-JDBC-Treiber, Version 1.2.27.1051

## Änderungen, Verbesserungen und behobene Probleme

- Spark
  - Es wurde eine Optimierung für das dynamische Bereinigen von Partitionen hinzugefügt. Die Optimierung ist standardmäßig deaktiviert. Um es zu aktivieren, setzen Sie den Spark-Konfigurationsparameter `spark.sql.dynamicPartitionPruning.enabled` auf `true`.
  - Die Leistung von INTERSECT-Anfragen wurde verbessert. Diese Optimierung ist standardmäßig deaktiviert. Um es zu aktivieren, setzen Sie den Spark-Konfigurationsparameter `spark.sql.optimizer.distinctBeforeIntersect.enabled` auf `true`.
  - Es wurde eine Optimierung hinzugefügt, um skalare Unterabfragen mit Aggregaten, die dieselbe Beziehung verwenden, zu vereinfachen. Die Optimierung ist standardmäßig deaktiviert. Um es zu aktivieren, setzen Sie den Spark-Konfigurationsparameter `spark.sql.optimizer.flattenScalarSubqueriesWithAggregates.enabled` auf `true`.
  - Die Codegenerierung für die gesamte Phase wurde verbessert.

Weitere Informationen finden Sie unter [Optimieren der Spark-Leistung](#).

- Verbesserungen der lokalen Festplattenverschlüsselung und der S3-Verschlüsselung in Sicherheitskonfigurationen (5. August 2019)
  - Die Amazon-S3-Verschlüsselungseinstellungen wurden im Sicherheitskonfigurations-Setup von den Einstellungen für die lokale Festplattenverschlüsselung getrennt.
  - Es wurde eine Option hinzugefügt, um die EBS-Verschlüsselung zu aktivieren. Bei Auswahl dieser Option wird zusätzlich zu den Speichervolumen auch das Root-Geräte-Volumen

verschlüsselt. Frühere Versionen erforderten die Verwendung eines benutzerdefinierten AMI, um das Root-Geräte-Volume zu verschlüsseln.

- Weitere Informationen erhalten Sie unter [Verschlüsselungsoptionen](#) im Verwaltungshandbuch für Amazon EMR.

## Bekannte Probleme

- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in Amazon-EMR-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie dem Herunterskalieren oder der schrittweisen Übermittlung auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

### Workaround:

- SSH als hadoop-Benutzer für den führenden Primärknoten des EMR-Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den hadoop-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

### Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

## 5.24.0 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.2.1	Amazon SageMaker Spark-SDK
<code>emr-ddb</code>	4.9.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.8.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-s3-dist-cp</code>	2.11.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
<code>emr-s3-select</code>	1.3.0	EMR S3Select-Konnektor

Komponente	Version	Beschreibung
emrfs	2.33.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.8.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.5-amzn-4	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.5-amzn-4	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.5-amzn-4	HDFS-Client und -Bibliothek für die Befehlszeile

Komponente	Version	Beschreibung
hadoop-hdfs-namenode	2.8.5-amzn-4	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-hdfs-journalnode	2.8.5-amzn-4	HDFS-Service zum Verwalten des Hadoop-Dateisystem journals auf HA-Clustern.
hadoop-httpfs-server	2.8.5-amzn-4	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.8.5-amzn-4	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.8.5-amzn-4	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.8.5-amzn-4	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resource-manager	2.8.5-amzn-4	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.5-amzn-4	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.

Komponente	Version	Beschreibung
hbase-hmaster	1.4.9	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.4.9	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.4.9	HBase-Befehlszeilen-Client.
hbase-rest-server	1.4.9	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.4.9	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.4-amzn-2	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.4-amzn-2	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.3.4-amzn-2	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.3.4-amzn-2	Hive-Befehlszeilen-Client.
hive-hbase	2.3.4-amzn-2	Hive-hbase client.

Komponente	Version	Beschreibung
hive-metastore-server	2.3.4-amzn-2	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.3.4-amzn-2	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.4.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	0.9.6	Multi-User-Server für Jupyter-Notebooks
livy-server	0.6.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP- und Reverse-Proxy-Server.
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.4.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Toolkit
oozie-client	5.1.0	Oozie-Befehlszeilen-Client.



Komponente	Version	Beschreibung
oozie-server	5.1.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.1-HBase-1.4	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.14.1-HBase-1.4	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.219	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.219	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.1	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
spark-client	2.4.2	Spark-Befehlszeilen-Clients.

Komponente	Version	Beschreibung
spark-history-server	2.4.2	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.4.2	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.4.2	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	1.12.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.1	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.8.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.13	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.

Komponente	Version	Beschreibung
zookeeper-client	3.4.13	ZooKeeper Befehlszeilen-Client.

## 5.24.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

### emr-5.24.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-log4j	Ändert die Werte in der <code>container-log4j.properties</code> -Datei in Hadoop YARN.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Ändert die <code>log4j-cli.properties</code> -Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.

Klassifizierungen	Beschreibung
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site

Klassifizierungen	Beschreibung
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.

Klassifizierungen	Beschreibung
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.

Klassifizierungen	Beschreibung
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.

Klassifizierungen	Beschreibung
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.



Klassifizierungen	Beschreibung
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.

Klassifizierungen	Beschreibung
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.23.1

### 5.23.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

## Informationen zur Anwendungsversion

	emr-5.23.1	emr-5.23.0	emr-5.22.0	emr-5.21.2
AWS SDK for Java	1.11.519	1,11.519	1,11.510	1.11.479
Python	2,7, 3,6	2,7, 3,6	2,7, 3,6	2,7, 3,6
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.7.1	1.7.1	1.7.1	1.7.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.9	1.4.9	1.4.9	1.4.8
HCatalog	2.3.4	2.3.4	2.3.4	2.3.4
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.4	2.3.4	2.3.4	2.3.4
Hudi	-	-	-	-
Hue	4.3.0	4.3.0	4.3.0	4.3.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.9.4	0.9.4	0.9.4	0.9.4
Livy	0.5.0	0.5.0	0.5.0	0.5.0
MXNet	1.3.1	1.3.1	1.3.1	1.3.1

	emr-5.23.1	emr-5.23.0	emr-5.22.0	emr-5.21.2
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.1.0	5.1.0	5.1.0	5.0.0
Phoenix	4.14.1	4.14.1	4.14.1	4.14.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.215	0.215	0.215	0.215
Spark	2.4.0	2.4.0	2.4.0	2.4.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.12.0	1.12.0	1.12.0	1.12.0
Tez	0.9.1	0.9.1	0.9.1	0.9.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.1	0.8.1	0.8.1	0.8.0
ZooKeeper	3.4.13	3.4.13	3.4.13	3.4.13

### 5.23.1 Versionshinweise

Dies ist eine Patch-Version. Alle Anwendungen und Komponenten sind dieselben wie in der vorherigen Amazon-EMR-Version.

Supportstatus für Instance Metadata Service (IMDS) V2: Amazon-EMR-Komponenten von 5.23.1, 5.27.1 und 5.32 oder höher verwenden IMDSv2 für alle IMDS-Aufrufe. Für IMDS-Aufrufe in Ihrem Anwendungscode können Sie sowohl IMDSv1 als auch IMDSv2 verwenden oder das IMDS so konfigurieren, dass es aus Sicherheitsgründen nur IMDSv2 verwendet. Bei anderen 5.x-EMR-Versionen führt die Deaktivierung von IMDSv1 zu einem Cluster-Startup-Fehler.

## 5.23.1 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.2.1	Amazon SageMaker Spark-SDK
<code>emr-ddb</code>	4.8.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.7.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-s3-dist-cp</code>	2.11.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
<code>emr-s3-select</code>	1.2.0	EMR S3Select-Konnektor

Komponente	Version	Beschreibung
emrfs	2.32.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.7.1	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.5-amzn-3	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.5-amzn-3	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.5-amzn-3	HDFS-Client und -Bibliothek für die Befehlszeile

Komponente	Version	Beschreibung
hadoop-hdfs-namenode	2.8.5-amzn-3	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-hdfs-journalnode	2.8.5-amzn-3	HDFS-Service zum Verwalten des Hadoop-Dateisystem journals auf HA-Clustern.
hadoop-httfs-server	2.8.5-amzn-3	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.8.5-amzn-3	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.8.5-amzn-3	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.8.5-amzn-3	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.8.5-amzn-3	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.5-amzn-3	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.

Komponente	Version	Beschreibung
hbase-hmaster	1.4.9	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.4.9	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.4.9	HBase-Befehlszeilen-Client.
hbase-rest-server	1.4.9	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.4.9	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.4-amzn-1	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.4-amzn-1	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.3.4-amzn-1	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.3.4-amzn-1	Hive-Befehlszeilen-Client.
hive-hbase	2.3.4-amzn-1	Hive-hbase client.



Komponente	Version	Beschreibung
hive-metastore-server	2.3.4-amzn-1	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.3.4-amzn-1	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.3.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	0.9.4	Multi-User-Server für Jupyter-Notebooks
livy-server	0.5.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP- und Reverse-Proxy-Server.
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.3.1	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Toolkit
oozie-client	5.1.0	Oozie-Befehlszeilen-Client.

Komponente	Version	Beschreibung
oozie-server	5.1.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.1-HBase-1.4	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.14.1-HBase-1.4	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.215	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.215	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.1	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
spark-client	2.4.0	Spark-Befehlszeilen-Clients.

Komponente	Version	Beschreibung
spark-history-server	2.4.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.4.0	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.4.0	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	1.12.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.1	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.8.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.13	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.

Komponente	Version	Beschreibung
zookeeper-client	3.4.13	ZooKeeper Befehlszeilen-Client.

### 5.23.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

#### emr-5.23.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-log4j	Ändert die Werte in der <code>container-log4j.properties</code> -Datei in Hadoop YARN.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Ändert die <code>log4j-cli.properties</code> -Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.

Klassifizierungen	Beschreibung
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site

Klassifizierungen	Beschreibung
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.

Klassifizierungen	Beschreibung
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.

Klassifizierungen	Beschreibung
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.



Klassifizierungen	Beschreibung
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.

Klassifizierungen	Beschreibung
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.23.0

### 5.23.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

## Informationen zur Anwendungsversion

	emr-5.23.0	emr-5.22.0	emr-5.21.2	emr-5.21.1
AWS SDK for Java	1.11.519	1,11.510	1.11.479	1.11.479
Python	2,7, 3,6	2,7, 3,6	2,7, 3,6	2,7, 3,6
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.7.1	1.7.1	1.7.0	1.7.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.9	1.4.9	1.4.8	1.4.8
HCatalog	2.3.4	2.3.4	2.3.4	2.3.4
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.4	2.3.4	2.3.4	2.3.4
Hudi	-	-	-	-
Hue	4.3.0	4.3.0	4.3.0	4.3.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.9.4	0.9.4	0.9.4	0.9.4
Livy	0.5.0	0.5.0	0.5.0	0.5.0
MXNet	1.3.1	1.3.1	1.3.1	1.3.1

	emr-5.23.0	emr-5.22.0	emr-5.21.2	emr-5.21.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.1.0	5.1.0	5.0.0	5.0.0
Phoenix	4.14.1	4.14.1	4.14.0	4.14.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.215	0.215	0.215	0.215
Spark	2.4.0	2.4.0	2.4.0	2.4.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.12.0	1.12.0	1.12.0	1.12.0
Tez	0.9.1	0.9.1	0.9.1	0.9.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.1	0.8.1	0.8.0	0.8.0
ZooKeeper	3.4.13	3.4.13	3.4.13	3.4.13

## 5.23.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.23.0. Änderungen beziehen sich auf Version 5.22.0.

Erste Version: 01. April 2019

Letzte Aktualisierung: 30. April 2019

### Upgrades

- AWS SDK for Java 1,11.519

## Neue Features

- (30. April 2019) Mit Amazon EMR 5.23.0 und höher können Sie einen Cluster mit drei primären Knoten starten, um die Hochverfügbarkeit von Anwendungen wie YARN Resource Manager, HDFS, Spark NameNode, Hive und Ganglia zu unterstützen. Der Primärknoten ist mit diesem Feature keine potenzielle einzelne Fehlerquelle mehr. Wenn ein Primärknoten ausfällt, führt Amazon EMR automatisch einen Failover zu einem Standby-Primärknoten aus und ersetzt den ausgefallenen Primärknoten durch einen neuen Primärknoten mit der gleichen Konfiguration und den gleichen Bootstrap-Aktionen. Weitere Informationen finden Sie unter [Primärknoten planen und konfigurieren](#).

## Bekannte Probleme

- Tez UI (In Amazon-EMR-Version 5.26.0 behoben)

Die Tez-Benutzeroberfläche funktioniert nicht auf einem EMR-Cluster mit mehreren Primärknoten.

- Hue (in Amazon-EMR-Version 5.24.0 behoben)
  - Hue, das auf Amazon EMR läuft, unterstützt Solr nicht. Ab Amazon-EMR-Version 5.20.0 führt ein Problem mit der Fehlkonfiguration dazu, dass Solr aktiviert wird und eine harmlose Fehlermeldung ähnlich der folgenden angezeigt wird:

```
Solr server could not be contacted properly:
HTTPConnectionPool('host=ip-xx-xx-xx-xx.ec2.internal',
port=1978): Max retries exceeded with url: /solr/admin/info/
system?user.name=hue&doAs=administrator&wt=json (Caused by
NewConnectionError(': Failed to establish a new connection: [Errno 111]
Connection refused',))
```

So wird verhindert, dass die Solr-Fehlermeldung angezeigt wird:

1. Stellen Sie über SSH eine Verbindung zur Befehlszeile des Primärknotens her.
2. Verwenden Sie einen Texteditor zum Öffnen der `hue.ini`-Datei. Beispielsweise:

```
sudo vim /etc/hue/conf/hue.ini
```

3. Suchen Sie nach dem Begriff `appblacklist` und ändern Sie die Zeile wie folgt:

```
appblacklist = search
```

4. Speichern Sie Ihre Änderungen und starten Sie Hue wie im folgenden Beispiel gezeigt:

```
sudo stop hue; sudo start hue
```

- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in Amazon-EMR-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie dem Herunterskalieren oder der schrittweisen Übermittlung auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

Workaround:

- SSH als hadoop-Benutzer für den führenden Primärknoten des EMR-Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den hadoop-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

#### Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

## 5.23.0 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version

sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-amzn-*EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen myapp-component der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als 2.2-amzn-2 aufgeführt.

Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1.2.1	Amazon SageMaker Spark-SDK
emr-ddb	4.8.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.7.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.11.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	1.2.0	EMR S3Select-Konnektor
emrfs	2.32.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.7.1	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.



Komponente	Version	Beschreibung
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.5-amzn-3	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.5-amzn-3	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.5-amzn-3	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.8.5-amzn-3	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-hdfs-journalnode	2.8.5-amzn-3	HDFS-Service zum Verwalten des Hadoop-Dateisystem journals auf HA-Clustern.

Komponente	Version	Beschreibung
hadoop-https-server	2.8.5-amzn-3	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.8.5-amzn-3	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.8.5-amzn-3	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.8.5-amzn-3	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resource-manager	2.8.5-amzn-3	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.5-amzn-3	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.4.9	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.4.9	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.

Komponente	Version	Beschreibung
hbase-client	1.4.9	HBase-Befehlszeilen-Client.
hbase-rest-server	1.4.9	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.4.9	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.4-amzn-1	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.4-amzn-1	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.3.4-amzn-1	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.3.4-amzn-1	Hive-Befehlszeilen-Client.
hive-hbase	2.3.4-amzn-1	Hive-hbase client.
hive-metastore-server	2.3.4-amzn-1	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.3.4-amzn-1	Service zur Annahme von Hive-Abfragen als Webanfragen.

Komponente	Version	Beschreibung
hue-server	4.3.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	0.9.4	Multi-User-Server für Jupyter-Notebooks
livy-server	0.5.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP- und Reverse-Proxy-Server.
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.3.1	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Toolkit
oozie-client	5.1.0	Oozie-Befehlszeilen-Client.
oozie-server	5.1.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.1-HBase-1.4	Die Phoenix-Bibliotheken für den Server und den Client

Komponente	Version	Beschreibung
phoenix-query-server	4.14.1-HBase-1.4	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.215	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.215	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.1	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
spark-client	2.4.0	Spark-Befehlszeilen-Clients.
spark-history-server	2.4.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.4.0	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.4.0	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.

Komponente	Version	Beschreibung
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	1.12.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.1	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.8.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.13	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.13	ZooKeeper Befehlszeilen-Client.

### 5.23.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

## emr-5.23.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der capacity-scheduler.xml-Datei in Hadoop.
container-log4j	Ändert die Werte in der container-log4j.properties-Datei in Hadoop YARN.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4j-yarn-session.properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.

Klassifizierungen	Beschreibung
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.



Klassifizierungen	Beschreibung
hive-parquet-logging	Ändert die Werte in der <code>parquet-logging.properties</code> -Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von <code>hive-exec-log</code> Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von Hive. <code>llap-daemon-log</code>
hive-log4j2	Ändert die Werte in der <code>hive-log4j2.properties</code> -Datei in Hive.
hive-site	Ändert die Werte in der <code>hive-site.xml</code> -Datei in Hive.
hiveserver2-site	Ändert die Werte in der <code>hiveserver2-site.xml</code> -Datei von <code>Server2</code> in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der <code>httpfs-site.xml</code> -Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der <code>kms-acls.xml</code> -Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der <code>kms-log4j.properties</code> -Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der <code>kms-site.xml</code> -Datei in Hadoop.

Klassifizierungen	Beschreibung
jupyter-notebook-conf	Ändert die Werte in der <code>jupyter_notebook_conf.py</code> -Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei <code>jupyterhub_config.py</code> .
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der <code>config.json</code> -Datei in Sparkmagic.
livy-conf	Ändert die Werte in der <code>livy.conf</code> -Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei <code>mapred-site.xml</code> der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der <code>oozie-log4j.properties</code> -Datei in Oozie.
oozie-site	Ändert die Werte in der <code>oozie-site.xml</code> -Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der <code>hadoop-metrics2-hbase.properties</code> -Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der <code>hbase-site.xml</code> -Datei in Phoenix.

Klassifizierungen	Beschreibung
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.

Klassifizierungen	Beschreibung
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.

Klassifizierungen	Beschreibung
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.22.0

### 5.22.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

### Informationen zur Anwendungsversion

	emr-5.22.0	emr-5.21.2	emr-5.21.1	emr-5.21.0
AWS SDK for Java	1.11.510	1.11.479	1.11.479	1.11.479
Python	2,7, 3,6	2,7, 3,6	2,7, 3,6	2,7, 3,6
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-

	emr-5.22.0	emr-5.21.2	emr-5.21.1	emr-5.21.0
Delta	-	-	-	-
Flink	1.7.1	1.7.0	1.7.0	1.7.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.9	1.4.8	1.4.8	1.4.8
HCatalog	2.3.4	2.3.4	2.3.4	2.3.4
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.4	2.3.4	2.3.4	2.3.4
Hudi	-	-	-	-
Hue	4.3.0	4.3.0	4.3.0	4.3.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.9.4	0.9.4	0.9.4	0.9.4
Livy	0.5.0	0.5.0	0.5.0	0.5.0
MXNet	1.3.1	1.3.1	1.3.1	1.3.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.1.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.1	4.14.0	4.14.0	4.14.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.215	0.215	0.215	0.215
Spark	2.4.0	2.4.0	2.4.0	2.4.0

	emr-5.22.0	emr-5.21.2	emr-5.21.1	emr-5.21.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.12.0	1.12.0	1.12.0	1.12.0
Tez	0.9.1	0.9.1	0.9.1	0.9.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.1	0.8.0	0.8.0	0.8.0
ZooKeeper	3.4.13	3.4.13	3.4.13	3.4.13

## 5.22.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.22.0. Änderungen beziehen sich auf Version 5.21.0.

### Important

Ab Amazon EMR Version 5.22.0 verwendet Amazon EMR AWS Signature Version 4 ausschließlich zur Authentifizierung von Anfragen an Amazon S3. Frühere Amazon EMR-Versionen verwenden in einigen Fällen AWS Signature Version 2, sofern in den Versionshinweisen nicht angegeben ist, dass ausschließlich Signature Version 4 verwendet wird. Weitere Informationen finden Sie unter [Authentifizieren von Anfragen \(AWS Signature Version 4\)](#) und [Authentifizieren von Anfragen \(AWS Signature Version 2\)](#) im Amazon Simple Storage Service Developer Guide.

Erste Version: 20. März 2019

### Upgrades

- Flink 1.7.1
- HBase 1.4.9
- Oozie 5.1.0



- Phoenix 4.14.1
- Zeppelin 0.8.1
- Konnektoren und Treiber:
  - DynamoDB Connector 4.8.0
  - MariaDB Connector 2.2.6
  - Amazon-Redshift-JDBC-Treiber, Version 1.2.20.1043

## Neue Features

- Die EBS-Standardkonfiguration für EC2-Instance-Typen mit reinem EBS-Speicher wurde geändert. Beim Erstellen eines Clusters mit Amazon-EMR-Version ab 5.22.0 steigt die Größe des standardmäßig zugewiesenen EBS-Speichers basierend auf der Größe der Instance. Darüber hinaus haben wir den erhöhten Speicher auf mehrere Volumes aufgeteilt, was zu einer höheren IOPS-Leistung führt. Wenn Sie eine andere EBS-Instance-Speicherkonfiguration verwenden möchten, können Sie diese beim Erstellen eines EMR-Clusters bzw. beim Hinzufügen von Knoten zu einem Cluster angeben. Weitere Informationen zur Speichermenge und Anzahl der standardmäßig zugewiesenen Volumes für jeden Instance-Typ finden Sie unter [Standard-EBS-Speicher für Instances](#) im Verwaltungshandbuch für Amazon EMR.

## Änderungen, Verbesserungen und behobene Probleme

- Spark
  - Es wurde eine neue Konfigurationseigenschaft für Spark auf YARN `spark.yarn.executor.memoryOverheadFactor` eingeführt. Der Wert dieser Eigenschaft ist ein Skalierungsfaktor, der den Wert des Speicher-Overheads auf einen Prozentsatz des Executor-Speichers festlegt, mit einem Minimum von 384 MB. Wenn der Speicher-Overhead explizit mit `spark.yarn.executor.memoryOverhead` festgelegt wird, hat diese Eigenschaft keine Auswirkung. Der Standardwert ist `0.1875`, was 18,75 % entspricht. Dieser Standard für Amazon EMR lässt mehr Platz in YARN-Containern für den Executor-Speicheraufwand übrig als der intern von Spark festgelegte Standard von 10 %. Der Amazon-EMR-Standard von 18,75 % ergab empirisch weniger speicherbedingte Ausfälle in TPC-DS-Benchmarks.
  - [SPARK-26316](#) wurde zurückportiert, um die Leistung zu verbessern.
- In den Amazon-EMR-Versionen 5.19.0, 5.20.0 und 5.21.0 werden YARN-Knotenbezeichnungen in einem HDFS-Verzeichnis gespeichert. In einigen Situationen führt dies zu Verzögerungen beim Startup des Core-Knotens und dann zu einem Cluster-Timeout und einem Startfehler. Ab Amazon

EMR 5.22.0 ist dieses Problem behoben. YARN-Knotenbezeichnungen werden auf der lokalen Festplatte jedes Clusterknotens gespeichert, wodurch Abhängigkeiten von HDFS vermieden werden.

## Bekannte Probleme

- Hue (in Amazon-EMR-Version 5.24.0 behoben)
  - Hue, das auf Amazon EMR läuft, unterstützt Solr nicht. Ab Amazon-EMR-Version 5.20.0 führt ein Problem mit der Fehlkonfiguration dazu, dass Solr aktiviert wird und eine harmlose Fehlermeldung ähnlich der folgenden angezeigt wird:

```
Solr server could not be contacted properly:
HTTPConnectionPool('host=ip-xx-xx-xx-xx.ec2.internal',
port=1978): Max retries exceeded with url: /solr/admin/info/
system?user.name=hue&doAs=administrator&wt=json (Caused by
NewConnectionError(': Failed to establish a new connection: [Errno 111]
Connection refused',))
```

So wird verhindert, dass die Solr-Fehlermeldung angezeigt wird:

1. Stellen Sie über SSH eine Verbindung zur Befehlszeile des Primärknotens her.
2. Verwenden Sie einen Texteditor zum Öffnen der `hue.ini`-Datei. Beispielsweise:

```
sudo vim /etc/hue/conf/hue.ini
```

3. Suchen Sie nach dem Begriff `appblacklist` und ändern Sie die Zeile wie folgt:

```
appblacklist = search
```

4. Speichern Sie Ihre Änderungen und starten Sie Hue wie im folgenden Beispiel gezeigt:

```
sudo stop hue; sudo start hue
```

- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in Amazon-EMR-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie dem Herunterskalieren oder der schrittweisen Übermittlung auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische

als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

Workaround:

- SSH als hadoop-Benutzer für den führenden Primärknoten des EMR-Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den hadoop-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

#### Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

## 5.22.0 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1.2.1	Amazon SageMaker Spark-SDK
emr-ddb	4.8.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.6.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.11.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	1.2.0	EMR S3Select-Konnektor
emrfs	2.31.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.7.1	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken

Komponente	Version	Beschreibung
		aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.5-amzn-2	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.5-amzn-2	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.5-amzn-2	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.8.5-amzn-2	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httpfs-server	2.8.5-amzn-2	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.8.5-amzn-2	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.8.5-amzn-2	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.

Komponente	Version	Beschreibung
hadoop-yarn-nodemanager	2.8.5-amzn-2	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.8.5-amzn-2	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.5-amzn-2	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.4.9	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.4.9	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.4.9	HBase-Befehlszeilen-Client.
hbase-rest-server	1.4.9	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.4.9	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.4-amzn-1	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.

Komponente	Version	Beschreibung
hcatalog-server	2.3.4-amzn-1	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.3.4-amzn-1	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.3.4-amzn-1	Hive-Befehlszeilen-Client.
hive-hbase	2.3.4-amzn-1	Hive-hbase client.
hive-metastore-server	2.3.4-amzn-1	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.3.4-amzn-1	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.3.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	0.9.4	Multi-User-Server für Jupyter-Notebooks
livy-server	0.5.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP- und Reverse-Proxy-Server.

Komponente	Version	Beschreibung
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.3.1	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.1.0	Oozie-Befehlszeilen-Client.
oozie-server	5.1.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.1-HBase-1.4	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.14.1-HBase-1.4	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.215	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.215	Service für das Ausführen von Teilen einer Abfrage.



Komponente	Version	Beschreibung
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.1	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
spark-client	2.4.0	Spark-Befehlszeilen-Clients.
spark-history-server	2.4.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.4.0	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.4.0	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	1.12.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.1	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.

Komponente	Version	Beschreibung
zeppelin-server	0.8.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.13	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.13	ZooKeeper Befehlszeilen-Cli-ent.

## 5.22.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

### emr-5.22.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-log4j	Ändert die Werte in der <code>container-log4j.properties</code> -Datei in Hadoop YARN.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.

Klassifizierungen	Beschreibung
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4 j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.

Klassifizierungen	Beschreibung
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.

Klassifizierungen	Beschreibung
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.

Klassifizierungen	Beschreibung
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.

Klassifizierungen	Beschreibung
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.



Klassifizierungen	Beschreibung
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.21.2

### 5.21.2 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

### Informationen zur Anwendungsversion

	emr-5.21.2	emr-5.21.1	emr-5.21.0	emr-5.20.1
AWS SDK for Java	1.11.479	1.11.479	1.11.479	1.11.461
Python	2,7, 3,6	2,7, 3,6	2,7, 3,6	2,7, 3,6
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.7.0	1.7.0	1.7.0	1.6.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.8	1.4.8	1.4.8	1.4.8
HCatalog	2.3.4	2.3.4	2.3.4	2.3.4
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.4	2.3.4	2.3.4	2.3.4
Hudi	-	-	-	-
Hue	4.3.0	4.3.0	4.3.0	4.3.0
Iceberg	-	-	-	-

	emr-5.21.2	emr-5.21.1	emr-5.21.0	emr-5.20.1
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.9.4	0.9.4	0.9.4	0.9.4
Livy	0.5.0	0.5.0	0.5.0	0.5.0
MXNet	1.3.1	1.3.1	1.3.1	1.3.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.0	4.14.0	4.14.0	4.14.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.215	0.215	0.215	0.214
Spark	2.4.0	2.4.0	2.4.0	2.4.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.12.0	1.12.0	1.12.0	1.12.0
Tez	0.9.1	0.9.1	0.9.1	0.9.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.0	0.8.0	0.8.0	0.8.0
ZooKeeper	3.4.13	3.4.13	3.4.13	3.4.13

## 5.21.2 Versionshinweise

Dies ist eine Patch-Version, um die AWS Signature Version 4-Authentifizierung für Anfragen an Amazon S3 hinzuzufügen. Alle Anwendungen und Komponenten sind dieselben wie in der vorherigen Amazon-EMR-Version.

### ⚠ Important

In dieser Release-Version verwendet Amazon EMR AWS Signature Version 4 ausschließlich zur Authentifizierung von Anfragen an Amazon S3. Weitere Informationen finden Sie unter [Was ist neu](#).

## 5.21.2 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.2.1	Amazon SageMaker Spark-SDK
<code>emr-ddb</code>	4.7.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.5.1	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-s3-dist-cp	2.11.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	1.2.0	EMR S3Select-Konnektor
emrfs	2.30.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.7.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.5-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.5-amzn-1	HDFS-Service auf Knotenebene zum Speichern von Blöcken.

Komponente	Version	Beschreibung
hadoop-hdfs-library	2.8.5-amzn-1	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.8.5-amzn-1	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httfs-server	2.8.5-amzn-1	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.8.5-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.8.5-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.8.5-amzn-1	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.8.5-amzn-1	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.5-amzn-1	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.

Komponente	Version	Beschreibung
hbase-hmaster	1.4.8	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.4.8	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.4.8	HBase-Befehlszeilen-Client.
hbase-rest-server	1.4.8	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.4.8	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.4-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.4-amzn-0	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.3.4-amzn-0	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.3.4-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.3.4-amzn-0	Hive-hbase client.

Komponente	Version	Beschreibung
hive-metastore-server	2.3.4-amzn-0	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.3.4-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.3.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	0.9.4	Multi-User-Server für Jupyter-Notebooks
livy-server	0.5.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP- und Reverse-Proxy-Server.
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.3.1	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Toolkit
oozie-client	5.0.0	Oozie-Befehlszeilen-Client.



Komponente	Version	Beschreibung
oozie-server	5.0.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.0-HBase-1.4	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.14.0-HBase-1.4	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.215	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.215	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.1	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
spark-client	2.4.0	Spark-Befehlszeilen-Clients.

Komponente	Version	Beschreibung
spark-history-server	2.4.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.4.0	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.4.0	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	1.12.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.1	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.8.0	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.13	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.

Komponente	Version	Beschreibung
zookeeper-client	3.4.13	ZooKeeper Befehlszeilen-Client.

## 5.21.2 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

### emr-5.21.2-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-log4j	Ändert die Werte in der <code>container-log4j.properties</code> -Datei in Hadoop YARN.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Ändert die <code>log4j-cli.properties</code> -Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.

Klassifizierungen	Beschreibung
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site

Klassifizierungen	Beschreibung
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.

Klassifizierungen	Beschreibung
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.

Klassifizierungen	Beschreibung
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.

Klassifizierungen	Beschreibung
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.



Klassifizierungen	Beschreibung
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.

Klassifizierungen	Beschreibung
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.21.1

### 5.21.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

## Informationen zur Anwendungsversion

	emr-5.21.1	emr-5.21.0	emr-5.20.1	emr-5.20.0
AWS SDK for Java	1.11.479	1.11.479	1.11.461	1.11.461
Python	2,7, 3,6	2,7, 3,6	2,7, 3,6	2,7, 3,6
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.7.0	1.7.0	1.6.2	1.6.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.8	1.4.8	1.4.8	1.4.8
HCatalog	2.3.4	2.3.4	2.3.4	2.3.4
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.4	2.3.4	2.3.4	2.3.4
Hudi	-	-	-	-
Hue	4.3.0	4.3.0	4.3.0	4.3.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.9.4	0.9.4	0.9.4	0.9.4
Livy	0.5.0	0.5.0	0.5.0	0.5.0
MXNet	1.3.1	1.3.1	1.3.1	1.3.1

	emr-5.21.1	emr-5.21.0	emr-5.20.1	emr-5.20.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.0	4.14.0	4.14.0	4.14.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.215	0.215	0.214	0.214
Spark	2.4.0	2.4.0	2.4.0	2.4.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.12.0	1.12.0	1.12.0	1.12.0
Tez	0.9.1	0.9.1	0.9.1	0.9.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.0	0.8.0	0.8.0	0.8.0
ZooKeeper	3.4.13	3.4.13	3.4.13	3.4.13

## 5.21.1 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.21.1. Änderungen beziehen sich auf Version 5.21.0.

Erste Version: 18. Juli 2019

Änderungen, Verbesserungen und behobene Probleme

- Das standardmäßige Amazon Linux AMI für Amazon EMR wurde aktualisiert, sodass es wichtige Sicherheitsupdates für den Linux-Kernel enthält, darunter das „TCP SACK Denial of Service Issue“ ([AWS-2019-005](#)).

## Bekannte Probleme

- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in Amazon-EMR-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie dem Herunterskalieren oder der schrittweisen Übermittlung auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

Workaround:

- SSH als hadoop-Benutzer für den führenden Primärknoten des EMR-Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den hadoop-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

### Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

## 5.21.1 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version

sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-amzn-*EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen myapp-component der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als 2.2-amzn-2 aufgeführt.

Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1.2.1	Amazon SageMaker Spark-SDK
emr-ddb	4.7.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.5.1	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.11.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	1.2.0	EMR S3Select-Konnektor
emrfs	2.30.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.7.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.

Komponente	Version	Beschreibung
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.5-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.5-amzn-1	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.5-amzn-1	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.8.5-amzn-1	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httpfs-server	2.8.5-amzn-1	HTTP-Endpunkt für HDFS-Operationen.

Komponente	Version	Beschreibung
hadoop-kms-server	2.8.5-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.8.5-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.8.5-amzn-1	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.8.5-amzn-1	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.5-amzn-1	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.4.8	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.4.8	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.4.8	HBase-Befehlszeilen-Client.



Komponente	Version	Beschreibung
hbase-rest-server	1.4.8	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.4.8	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.4-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.4-amzn-0	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.3.4-amzn-0	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.3.4-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.3.4-amzn-0	Hive-hbase client.
hive-metastore-server	2.3.4-amzn-0	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.3.4-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.

Komponente	Version	Beschreibung
hue-server	4.3.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	0.9.4	Multi-User-Server für Jupyter-Notebooks
livy-server	0.5.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP- und Reverse-Proxy-Server.
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.3.1	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Toolkit
oozie-client	5.0.0	Oozie-Befehlszeilen-Client.
oozie-server	5.0.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.0-HBase-1.4	Die Phoenix-Bibliotheken für den Server und den Client

Komponente	Version	Beschreibung
phoenix-query-server	4.14.0-HBase-1.4	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.215	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.215	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.1	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
spark-client	2.4.0	Spark-Befehlszeilen-Clients.
spark-history-server	2.4.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.4.0	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.4.0	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.

Komponente	Version	Beschreibung
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	1.12.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.1	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.8.0	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.13	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.13	ZooKeeper Befehlszeilen-Client.

### 5.21.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

## emr-5.21.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der capacity-scheduler.xml-Datei in Hadoop.
container-log4j	Ändert die Werte in der container-log4j.properties-Datei in Hadoop YARN.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4j-yarn-session.properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.

Klassifizierungen	Beschreibung
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.

Klassifizierungen	Beschreibung
hive-parquet-logging	Ändert die Werte in der <code>parquet-logging.properties</code> -Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von <code>hive-exec-log</code> Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von Hive. <code>llap-daemon-log</code>
hive-log4j2	Ändert die Werte in der <code>hive-log4j2.properties</code> -Datei in Hive.
hive-site	Ändert die Werte in der <code>hive-site.xml</code> -Datei in Hive.
hiveserver2-site	Ändert die Werte in der <code>hiveserver2-site.xml</code> -Datei von <code>Server2</code> in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der <code>httpfs-site.xml</code> -Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der <code>kms-acls.xml</code> -Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der <code>kms-log4j.properties</code> -Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der <code>kms-site.xml</code> -Datei in Hadoop.

Klassifizierungen	Beschreibung
jupyter-notebook-conf	Ändert die Werte in der <code>jupyter_notebook_conf.py</code> -Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei <code>jupyterhub_config.py</code> .
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der <code>config.json</code> -Datei in Sparkmagic.
livy-conf	Ändert die Werte in der <code>livy.conf</code> -Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei <code>mapred-site.xml</code> der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der <code>oozie-log4j.properties</code> -Datei in Oozie.
oozie-site	Ändert die Werte in der <code>oozie-site.xml</code> -Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der <code>hadoop-metrics2-hbase.properties</code> -Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der <code>hbase-site.xml</code> -Datei in Phoenix.



Klassifizierungen	Beschreibung
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.

Klassifizierungen	Beschreibung
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.

Klassifizierungen	Beschreibung
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.21.0

### 5.21.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

### Informationen zur Anwendungsversion

	emr-5.21.0	emr-5.20.1	emr-5.20.0	emr-5.19.1
AWS SDK for Java	1.11.479	1.11.461	1.11.461	1,11.433
Python	2,7, 3,6	2,7, 3,6	2,7, 3,6	2,7, 3,4
Scala	2.11.12	2.11.12	2.11.12	2.11.8
AmazonCloudWatchAgent	-	-	-	-

	emr-5.21.0	emr-5.20.1	emr-5.20.0	emr-5.19.1
Delta	-	-	-	-
Flink	1.7.0	1.6.2	1.6.2	1.6.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.8	1.4.8	1.4.8	1.4.7
HCatalog	2.3.4	2.3.4	2.3.4	2.3.3
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.4	2.3.4	2.3.4	2.3.3
Hudi	-	-	-	-
Hue	4.3.0	4.3.0	4.3.0	4.2.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.9.4	0.9.4	0.9.4	0.9.4
Livy	0.5.0	0.5.0	0.5.0	0.5.0
MXNet	1.3.1	1.3.1	1.3.1	1.3.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.0	4.14.0	4.14.0	4.14.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.215	0.214	0.214	0.212
Spark	2.4.0	2.4.0	2.4.0	2.3.2

	emr-5.21.0	emr-5.20.1	emr-5.20.0	emr-5.19.1
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.12.0	1.12.0	1.12.0	1.11.0
Tez	0.9.1	0.9.1	0.9.1	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.0	0.8.0	0.8.0	0.8.0
ZooKeeper	3.4.13	3.4.13	3.4.13	3.4.13

## 5.21.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.21.0. Änderungen beziehen sich auf Version 5.20.0.

Veröffentlichungsdatum: 18. Februar 2019

Letzte Aktualisierung: 3. April 2019

### Upgrades

- Flink 1.7.0
- Presto 0.215
- AWS SDK for Java 1,11.479

### Neue Features

- (3. April 2019) Ab Amazon-EMR-Version 5.21.0 können Sie Cluster-Konfigurationen überschreiben und zusätzliche Konfigurationsklassifikationen für jede Instance-Gruppe in einem ausgeführten Cluster angeben. Dazu verwenden Sie die Amazon EMR-Konsole, das AWS Command Line Interface (AWS CLI) oder das AWS SDK. Weitere Informationen finden Sie unter [Angeben einer Konfiguration für eine Instance-Gruppe in einem aktiven Cluster](#).

## Änderungen, Verbesserungen und behobene Probleme

- Zeppelin
  - [ZEPPELIN-3878](#) rückportiert.

## Bekannte Probleme

- Hue (in Amazon-EMR-Version 5.24.0 behoben)
  - Hue, das auf Amazon EMR läuft, unterstützt Solr nicht. Ab Amazon-EMR-Version 5.20.0 führt ein Problem mit der Fehlkonfiguration dazu, dass Solr aktiviert wird und eine harmlose Fehlermeldung ähnlich der folgenden angezeigt wird:

```
Solr server could not be contacted properly:
HTTPConnectionPool('host=ip-xx-xx-xx-xx.ec2.internal',
port=1978): Max retries exceeded with url: /solr/admin/info/
system?user.name=hue&doAs=administrator&wt=json (Caused by
NewConnectionError(': Failed to establish a new connection: [Errno 111]
Connection refused',))
```

So wird verhindert, dass die Solr-Fehlermeldung angezeigt wird:

1. Stellen Sie über SSH eine Verbindung zur Befehlszeile des Primärknotens her.
2. Verwenden Sie einen Texteditor zum Öffnen der hue.ini-Datei. Beispielsweise:

```
sudo vim /etc/hue/conf/hue.ini
```

3. Suchen Sie nach dem Begriff `appblacklist` und ändern Sie die Zeile wie folgt:

```
appblacklist = search
```

4. Speichern Sie Ihre Änderungen und starten Sie Hue wie im folgenden Beispiel gezeigt:

```
sudo stop hue; sudo start hue
```

- Tez
  - Dieses Problem wurde in Amazon EMR 5.22.0 behoben.

Wenn Sie über eine SSH-Verbindung zum Primärknoten des Clusters eine Verbindung zur Tez-Benutzeroberfläche unter `http://MasterDNS:8080/tez-ui` herstellen, wird der Fehler `Adapterbetrieb fehlgeschlagen – Timeline-Server (ATS) ist nicht erreichbar` angezeigt.

Entweder ist er ausgefallen oder CORS ist nicht aktiviert“ wird angezeigt, oder Aufgaben zeigen unerwartet „N/A“ an.

Dies wird dadurch verursacht, dass die Tez-Benutzeroberfläche Anfragen an den YARN Timeline Server stellt und localhost nicht den Hostnamen des Primärknotens verwendet. Um dieses Problem zu umgehen, steht ein Skript zur Verfügung, das als Bootstrap-Aktion oder als Bootstrap-Schritt ausgeführt werden kann. Das Skript aktualisiert den Hostnamen in der `configs.env` Tez-Datei. Weitere Informationen und den Speicherort des Skripts finden Sie in den [Bootstrap-Anweisungen](#).

- In den Amazon-EMR-Versionen 5.19.0, 5.20.0 und 5.21.0 werden YARN-Knotenbezeichnungen in einem HDFS-Verzeichnis gespeichert. In einigen Situationen führt dies zu Verzögerungen beim Startup des Core-Knotens und dann zu einem Cluster-Timeout und einem Startfehler. Ab Amazon EMR 5.22.0 ist dieses Problem behoben. YARN-Knotenbezeichnungen werden auf der lokalen Festplatte jedes Clusterknotens gespeichert, wodurch Abhängigkeiten von HDFS vermieden werden.
- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in Amazon-EMR-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie dem Herunterskalieren oder der schrittweisen Übermittlung auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

Workaround:

- SSH als hadoop-Benutzer für den führenden Primärknoten des EMR-Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den hadoop-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.



### Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

## 5.21.0 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.2.1	Amazon SageMaker Spark-SDK
<code>emr-ddb</code>	4.7.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.5.1	Praktische Bibliotheken für das Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-kinesis	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.11.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	1.2.0	EMR S3Select-Konnektor
emrfs	2.30.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.7.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.5-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".

Komponente	Version	Beschreibung
hadoop-hdfs-datanode	2.8.5-amzn-1	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.5-amzn-1	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.8.5-amzn-1	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httpfs-server	2.8.5-amzn-1	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.8.5-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.8.5-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.8.5-amzn-1	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resource-manager	2.8.5-amzn-1	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.

Komponente	Version	Beschreibung
hadoop-yarn-timeline-server	2.8.5-amzn-1	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.4.8	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.4.8	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.4.8	HBase-Befehlszeilen-Client.
hbase-rest-server	1.4.8	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.4.8	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.4-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.4-amzn-0	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.3.4-amzn-0	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.

Komponente	Version	Beschreibung
hive-client	2.3.4-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.3.4-amzn-0	Hive-hbase client.
hive-metastore-server	2.3.4-amzn-0	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.3.4-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.3.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	0.9.4	Multi-User-Server für Jupyter-Notebooks
livy-server	0.5.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP- und Reverse-Proxy-Server.
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.3.1	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.

Komponente	Version	Beschreibung
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.0.0	Oozie-Befehlszeilen-Client.
oozie-server	5.0.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.0-HBase-1.4	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.14.0-HBase-1.4	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.215	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.215	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.1	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
spark-client	2.4.0	Spark-Befehlszeilen-Clients.

Komponente	Version	Beschreibung
spark-history-server	2.4.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.4.0	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.4.0	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	1.12.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.1	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.8.0	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.13	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.

Komponente	Version	Beschreibung
zookeeper-client	3.4.13	ZooKeeper Befehlszeilen-Client.

## 5.21.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

### emr-5.21.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-log4j	Ändert die Werte in der <code>container-log4j.properties</code> -Datei in Hadoop YARN.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Ändert die <code>log4j-cli.properties</code> -Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.



Klassifizierungen	Beschreibung
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site

Klassifizierungen	Beschreibung
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.

Klassifizierungen	Beschreibung
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.

Klassifizierungen	Beschreibung
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.

Klassifizierungen	Beschreibung
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.

Klassifizierungen	Beschreibung
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.20.1

### 5.20.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

## Informationen zur Anwendungsversion

	emr-5.20.1	emr-5.20.0	emr-5.19.1	emr-5.19.0
AWS SDK for Java	1.11.461	1.11.461	1,11.433	1,11.433
Python	2,7, 3,6	2,7, 3,6	2,7, 3,4	2,7, 3,4
Scala	2.11.12	2.11.12	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.6.2	1.6.2	1.6.1	1.6.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.8	1.4.8	1.4.7	1.4.7
HCatalog	2.3.4	2.3.4	2.3.3	2.3.3
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.4	2.3.4	2.3.3	2.3.3
Hudi	-	-	-	-
Hue	4.3.0	4.3.0	4.2.0	4.2.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.9.4	0.9.4	0.9.4	0.9.4
Livy	0.5.0	0.5.0	0.5.0	0.5.0
MXNet	1.3.1	1.3.1	1.3.0	1.3.0



	emr-5.20.1	emr-5.20.0	emr-5.19.1	emr-5.19.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.0	4.14.0	4.14.0	4.14.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.214	0.214	0.212	0.212
Spark	2.4.0	2.4.0	2.3.2	2.3.2
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.12.0	1.12.0	1.11.0	1.11.0
Tez	0.9.1	0.9.1	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.0	0.8.0	0.8.0	0.8.0
ZooKeeper	3.4.13	3.4.13	3.4.13	3.4.13

## 5.20.1 Versionshinweise

Dies ist eine Patch-Version, um die AWS Signature Version 4-Authentifizierung für Anfragen an Amazon S3 hinzuzufügen. Alle Anwendungen und Komponenten sind dieselben wie in der vorherigen Amazon-EMR-Version.

### Important

In dieser Release-Version verwendet Amazon EMR AWS Signature Version 4 ausschließlich zur Authentifizierung von Anfragen an Amazon S3. Weitere Informationen finden Sie unter [Was ist neu](#).

## 5.20.1 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.2.1	Amazon SageMaker Spark-SDK
<code>emr-ddb</code>	4.7.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.5.1	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-s3-dist-cp</code>	2.10.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
<code>emr-s3-select</code>	1.2.0	EMR S3Select-Konnektor

Komponente	Version	Beschreibung
emrfs	2.29.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.6.2	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.5-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.5-amzn-1	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.5-amzn-1	HDFS-Client und -Bibliothek für die Befehlszeile

Komponente	Version	Beschreibung
hadoop-hdfs-namenode	2.8.5-amzn-1	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httpfs-server	2.8.5-amzn-1	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.8.5-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.8.5-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.8.5-amzn-1	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.8.5-amzn-1	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.5-amzn-1	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.

Komponente	Version	Beschreibung
hbase-hmaster	1.4.8	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.4.8	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.4.8	HBase-Befehlszeilen-Client.
hbase-rest-server	1.4.8	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.4.8	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.4-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.4-amzn-0	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.3.4-amzn-0	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.3.4-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.3.4-amzn-0	Hive-hbase client.

Komponente	Version	Beschreibung
hive-metastore-server	2.3.4-amzn-0	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.3.4-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.3.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	0.9.4	Multi-User-Server für Jupyter-Notebooks
livy-server	0.5.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP- und Reverse-Proxy-Server.
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.3.1	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Toolkit
oozie-client	5.0.0	Oozie-Befehlszeilen-Client.

Komponente	Version	Beschreibung
oozie-server	5.0.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.0-HBase-1.4	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.14.0-HBase-1.4	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.214	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.214	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.1	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
spark-client	2.4.0	Spark-Befehlszeilen-Clients.

Komponente	Version	Beschreibung
spark-history-server	2.4.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.4.0	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.4.0	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	1.12.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.1	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.8.0	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.13	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.



Komponente	Version	Beschreibung
zookeeper-client	3.4.13	ZooKeeper Befehlszeilen-Client.

## 5.20.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

### emr-5.20.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-log4j	Ändert die Werte in der <code>container-log4j.properties</code> -Datei in Hadoop YARN.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Ändert die <code>log4j-cli.properties</code> -Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.

Klassifizierungen	Beschreibung
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site

Klassifizierungen	Beschreibung
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.

Klassifizierungen	Beschreibung
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.

Klassifizierungen	Beschreibung
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.

Klassifizierungen	Beschreibung
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.

Klassifizierungen	Beschreibung
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.20.0

### 5.20.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)



## Informationen zur Anwendungsversion

	emr-5.20.0	emr-5.19.1	emr-5.19.0	emr-5.18.1
AWS SDK for Java	1.11.461	1,11.433	1,11.433	1,11.393
Python	2,7, 3,6	2,7, 3,4	2,7, 3,4	2,7, 3,4
Scala	2.11.12	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.6.2	1.6.1	1.6.1	1.6.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.8	1.4.7	1.4.7	1.4.7
HCatalog	2.3.4	2.3.3	2.3.3	2.3.3
Hadoop	2.8.5	2.8.5	2.8.5	2.8.4
Hive	2.3.4	2.3.3	2.3.3	2.3.3
Hudi	-	-	-	-
Hue	4.3.0	4.2.0	4.2.0	4.2.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.9.4	0.9.4	0.9.4	0.8.1
Livy	0.5.0	0.5.0	0.5.0	0.5.0
MXNet	1.3.1	1.3.0	1.3.0	1.2.0

	emr-5.20.0	emr-5.19.1	emr-5.19.0	emr-5.18.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.0	4.14.0	4.14.0	4.14.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.214	0.212	0.212	0.210
Spark	2.4.0	2.3.2	2.3.2	2.3.2
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.12.0	1.11.0	1.11.0	1.9.0
Tez	0.9.1	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.0	0.8.0	0.8.0	0.8.0
ZooKeeper	3.4.13	3.4.13	3.4.13	3.4.12

## 5.20.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.20.0. Änderungen beziehen sich auf Version 5.19.0.

Veröffentlichungsdatum: 18. Dezember 2018

Letzte Aktualisierung: 22. Januar 2019

### Upgrades

- Flink 1.6.2
- HBase 1.4.8
- Hive 2.3.4

- Hue 4.3.0
- MXNet 1.3.1
- Presto 0.214
- Spark 2.4.0
- TensorFlow 1.12.0
- Tez 0.9.1
- AWS SDK for Java 1.11.461

## Neue Features

- (22. Januar 2019) Kerberos in Amazon EMR wurde verbessert und unterstützt nun die Authentifizierung von Prinzipalen von einem externen KDC. Dadurch wird die Prinzipalverwaltung zentralisiert, da mehrere Cluster einen einzelnen externen KDC gemeinsam nutzen können. Darüber hinaus kann der externe KDC eine bereichsübergreifende Vertrauensstellung mit einer Active-Directory-Domain haben. Auf diese Weise können alle Cluster Prinzipale von Active Directory authentifizieren. Weitere Informationen finden Sie unter [Verwendung der Kerberos-Authentifizierung](#) im Verwaltungshandbuch für Amazon EMR.

## Änderungen, Verbesserungen und behobene Probleme

- Standard-Amazon-Linux-AMI für Amazon EMR
  - Das Python3-Paket wurde von Python 3.4 auf 3.6 aktualisiert.
- Der EMRFS-S3-optimierte Committer
  - Der für EMRFS S3 optimierte Committer ist jetzt standardmäßig aktiviert, was die Schreibleistung verbessert. Weitere Informationen finden Sie unter [EMRFS-S3-optimierte Committer verwenden](#).
- Hive
  - [HIVE-16686](#) wurde rückportiert.
- Mit Spark und Hive Glue
  - In EMR 5.20.0 oder höher wird das parallel Partitionsbereinigen automatisch für Spark und Hive aktiviert, wenn AWS Glue Data Catalog als Metastore verwendet wird. Diese Änderung reduziert die Zeit für die Abfrageplanung erheblich, da mehrere Anfragen parallel ausgeführt werden, um Partitionen abzurufen. Die Gesamtzahl der Segmente, die gleichzeitig ausgeführt werden können, liegt zwischen 1 und 10. Der Standardwert ist 5, was eine empfohlene Einstellung ist. Sie können dies ändern, indem Sie die Eigenschaft `aws.glue.partition.num.segments` in

der `hive-site`-Konfigurationsklassifizierung angeben. Wenn eine Drosselung auftritt, können Sie das Feature ausschalten, indem Sie den Wert auf 1 ändern. Weitere Informationen finden Sie unter [AWS -Glue-Segmentstruktur](#).

## Bekannte Probleme

- Hue (in Amazon-EMR-Version 5.24.0 behoben)
  - Hue, das auf Amazon EMR läuft, unterstützt Solr nicht. Ab Amazon-EMR-Version 5.20.0 führt ein Problem mit der Fehlkonfiguration dazu, dass Solr aktiviert wird und eine harmlose Fehlermeldung ähnlich der folgenden angezeigt wird:

```
Solr server could not be contacted properly:
HTTPConnectionPool('host=ip-xx-xx-xx-xx.ec2.internal',
port=1978): Max retries exceeded with url: /solr/admin/info/
system?user.name=hue&doAs=administrator&wt=json (Caused by
NewConnectionError(': Failed to establish a new connection: [Errno 111]
Connection refused',))
```

So wird verhindert, dass die Solr-Fehlermeldung angezeigt wird:

1. Stellen Sie über SSH eine Verbindung zur Befehlszeile des Primärknotens her.
2. Verwenden Sie einen Texteditor zum Öffnen der `hue.ini`-Datei. Beispielsweise:

```
sudo vim /etc/hue/conf/hue.ini
```

3. Suchen Sie nach dem Begriff `appblacklist` und ändern Sie die Zeile wie folgt:

```
appblacklist = search
```

4. Speichern Sie Ihre Änderungen und starten Sie Hue wie im folgenden Beispiel gezeigt:

```
sudo stop hue; sudo start hue
```

- Tez
  - Dieses Problem wurde in Amazon EMR 5.22.0 behoben.

Wenn Sie über eine SSH-Verbindung zum Primärknoten des Clusters eine Verbindung zur Tez-Benutzeroberfläche unter `http://MasterDNS:8080/tez-ui` herstellen, wird der Fehler „Adapterbetrieb fehlgeschlagen – Timeline-Server (ATS) ist nicht erreichbar“ angezeigt.

Entweder ist er ausgefallen oder CORS ist nicht aktiviert“ wird angezeigt, oder Aufgaben zeigen unerwartet „N/A“ an.

Dies wird dadurch verursacht, dass die Tez-Benutzeroberfläche Anfragen an den YARN Timeline Server stellt und localhost nicht den Hostnamen des Primärknotens verwendet. Um dieses Problem zu umgehen, steht ein Skript zur Verfügung, das als Bootstrap-Aktion oder als Bootstrap-Schritt ausgeführt werden kann. Das Skript aktualisiert den Hostnamen in der `configs.env` Tez-Datei. Weitere Informationen und den Speicherort des Skripts finden Sie in den [Bootstrap-Anweisungen](#).

- In den Amazon-EMR-Versionen 5.19.0, 5.20.0 und 5.21.0 werden YARN-Knotenbezeichnungen in einem HDFS-Verzeichnis gespeichert. In einigen Situationen führt dies zu Verzögerungen beim Startup des Core-Knotens und dann zu einem Cluster-Timeout und einem Startfehler. Ab Amazon EMR 5.22.0 ist dieses Problem behoben. YARN-Knotenbezeichnungen werden auf der lokalen Festplatte jedes Clusterknotens gespeichert, wodurch Abhängigkeiten von HDFS vermieden werden.
- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in Amazon-EMR-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie dem Herunterskalieren oder der schrittweisen Übermittlung auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

Workaround:

- SSH als hadoop-Benutzer für den führenden Primärknoten des EMR-Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den hadoop-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

### Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

## 5.20.0 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.2.1	Amazon SageMaker Spark-SDK
<code>emr-ddb</code>	4.7.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.5.1	Praktische Bibliotheken für das Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-kinesis	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.10.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	1.2.0	EMR S3Select-Konnektor
emrfs	2.29.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.6.2	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.5-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".

Komponente	Version	Beschreibung
hadoop-hdfs-datanode	2.8.5-amzn-1	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.5-amzn-1	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.8.5-amzn-1	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httpfs-server	2.8.5-amzn-1	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.8.5-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.8.5-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.8.5-amzn-1	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resource-manager	2.8.5-amzn-1	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.



Komponente	Version	Beschreibung
hadoop-yarn-timeline-server	2.8.5-amzn-1	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.4.8	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.4.8	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.4.8	HBase-Befehlszeilen-Client.
hbase-rest-server	1.4.8	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.4.8	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.4-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.4-amzn-0	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.3.4-amzn-0	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.

Komponente	Version	Beschreibung
hive-client	2.3.4-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.3.4-amzn-0	Hive-hbase client.
hive-metastore-server	2.3.4-amzn-0	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.3.4-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.3.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	0.9.4	Multi-User-Server für Jupyter-Notebooks
livy-server	0.5.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP- und Reverse-Proxy-Server.
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.3.1	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.

Komponente	Version	Beschreibung
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.0.0	Oozie-Befehlszeilen-Client.
oozie-server	5.0.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.0-HBase-1.4	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.14.0-HBase-1.4	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.214	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.214	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.1	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
spark-client	2.4.0	Spark-Befehlszeilen-Clients.

Komponente	Version	Beschreibung
spark-history-server	2.4.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.4.0	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.4.0	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	1.12.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.9.1	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.8.0	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.13	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.

Komponente	Version	Beschreibung
zookeeper-client	3.4.13	ZooKeeper Befehlszeilen-Client.

## 5.20.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

### emr-5.20.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-log4j	Ändert die Werte in der <code>container-log4j.properties</code> -Datei in Hadoop YARN.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Ändert die <code>log4j-cli.properties</code> -Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.

Klassifizierungen	Beschreibung
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site

Klassifizierungen	Beschreibung
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.

Klassifizierungen	Beschreibung
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.



Klassifizierungen	Beschreibung
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.

Klassifizierungen	Beschreibung
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.

Klassifizierungen	Beschreibung
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.19.1

### 5.19.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

## Informationen zur Anwendungsversion

	emr-5.19.1	emr-5.19.0	emr-5.18.1	emr-5.18.0
AWS SDK for Java	1.11.433	1,11.433	1,11.393	1,11.393
Python	2,7, 3,4	2,7, 3,4	2,7, 3,4	2,7, 3,4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.6.1	1.6.1	1.6.0	1.6.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.7	1.4.7	1.4.7	1.4.7
HCatalog	2.3.3	2.3.3	2.3.3	2.3.3
Hadoop	2.8.5	2.8.5	2.8.4	2.8.4
Hive	2.3.3	2.3.3	2.3.3	2.3.3
Hudi	-	-	-	-
Hue	4.2.0	4.2.0	4.2.0	4.2.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.9.4	0.9.4	0.8.1	0.8.1
Livy	0.5.0	0.5.0	0.5.0	0.5.0
MXNet	1.3.0	1.3.0	1.2.0	1.2.0

	emr-5.19.1	emr-5.19.0	emr-5.18.1	emr-5.18.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.0	4.14.0	4.14.0	4.14.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.212	0.212	0.210	0.210
Spark	2.3.2	2.3.2	2.3.2	2.3.2
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.11.0	1.11.0	1.9.0	1.9.0
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.0	0.8.0	0.8.0	0.8.0
ZooKeeper	3.4.13	3.4.13	3.4.12	3.4.12

## 5.19.1 Versionshinweise

Dies ist eine Patch-Version, um die AWS Signature Version 4-Authentifizierung für Anfragen an Amazon S3 hinzuzufügen. Alle Anwendungen und Komponenten sind dieselben wie in der vorherigen Amazon-EMR-Version.

### Important

In dieser Release-Version verwendet Amazon EMR AWS Signature Version 4 ausschließlich zur Authentifizierung von Anfragen an Amazon S3. Weitere Informationen finden Sie unter [Was ist neu](#).

## 5.19.1 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.2.0	Amazon SageMaker Spark-SDK
<code>emr-ddb</code>	4.7.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.5.1	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-s3-dist-cp</code>	2.10.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
<code>emr-s3-select</code>	1.1.0	EMR S3Select-Konnektor

Komponente	Version	Beschreibung
emrfs	2.28.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.6.1	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.5-amzn-0	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.5-amzn-0	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.5-amzn-0	HDFS-Client und -Bibliothek für die Befehlszeile



Komponente	Version	Beschreibung
hadoop-hdfs-namenode	2.8.5-amzn-0	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httpfs-server	2.8.5-amzn-0	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.8.5-amzn-0	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.8.5-amzn-0	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.8.5-amzn-0	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.8.5-amzn-0	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.5-amzn-0	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.

Komponente	Version	Beschreibung
hbase-hmaster	1.4.7	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.4.7	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.4.7	HBase-Befehlszeilen-Client.
hbase-rest-server	1.4.7	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.4.7	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.3-amzn-2	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.3-amzn-2	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.3.3-amzn-2	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.3.3-amzn-2	Hive-Befehlszeilen-Client.
hive-hbase	2.3.3-amzn-2	Hive-hbase client.

Komponente	Version	Beschreibung
hive-metastore-server	2.3.3-amzn-2	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.3.3-amzn-2	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.2.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	0.9.4	Multi-User-Server für Jupyter-Notebooks
livy-server	0.5.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP- und Reverse-Proxy-Server.
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.3.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Toolkit
oozie-client	5.0.0	Oozie-Befehlszeilen-Client.

Komponente	Version	Beschreibung
oozie-server	5.0.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.0-HBase-1.4	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.14.0-HBase-1.4	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.212	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.212	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.1	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
spark-client	2.3.2	Spark-Befehlszeilen-Clients.

Komponente	Version	Beschreibung
spark-history-server	2.3.2	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.3.2	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.3.2	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	1.11.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.8.0	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.13	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.

Komponente	Version	Beschreibung
zookeeper-client	3.4.13	ZooKeeper Befehlszeilen-Client.

## 5.19.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

### emr-5.19.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-log4j	Ändert die Werte in der <code>container-log4j.properties</code> -Datei in Hadoop YARN.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Ändert die <code>log4j-cli.properties</code> -Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.

Klassifizierungen	Beschreibung
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site

Klassifizierungen	Beschreibung
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.



Klassifizierungen	Beschreibung
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.

Klassifizierungen	Beschreibung
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.

Klassifizierungen	Beschreibung
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.

Klassifizierungen	Beschreibung
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.19.0

### 5.19.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

## Informationen zur Anwendungsversion

	emr-5.19.0	emr-5.18.1	emr-5.18.0	emr-5.17.2
AWS SDK for Java	1.11.433	1,11.393	1,11.393	1.11.336
Python	2,7, 3,4	2,7, 3,4	2,7, 3,4	2,7, 3,4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.6.1	1.6.0	1.6.0	1.5.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.7	1.4.7	1.4.7	1.4.6
HCatalog	2.3.3	2.3.3	2.3.3	2.3.3
Hadoop	2.8.5	2.8.4	2.8.4	2.8.4
Hive	2.3.3	2.3.3	2.3.3	2.3.3
Hudi	-	-	-	-
Hue	4.2.0	4.2.0	4.2.0	4.2.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.9.4	0.8.1	0.8.1	0.8.1
Livy	0.5.0	0.5.0	0.5.0	0.5.0
MXNet	1.3.0	1.2.0	1.2.0	1.2.0

	emr-5.19.0	emr-5.18.1	emr-5.18.0	emr-5.17.2
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.0	4.14.0	4.14.0	4.14.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.212	0.210	0.210	0.206
Spark	2.3.2	2.3.2	2.3.2	2.3.1
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.11.0	1.9.0	1.9.0	1.9.0
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.0	0.8.0	0.8.0	0.7.3
ZooKeeper	3.4.13	3.4.12	3.4.12	3.4.12

## 5.19.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.19.0. Änderungen beziehen sich auf Version 5.18.0.

Erste Version: 7. November 2018

Letzte Aktualisierung: 19. November 2018

### Upgrades

- Hadoop 2.8.5
- Flink 1.6.1
- JupyterHub 0.9.4

- MXNet 1.3.0
- Presto 0.212
- TensorFlow 1.11.0
- Zookeeper 3.4.13
- AWS SDK for Java 1.11.433

## Neue Features

- (19. November 2018) EMR Notebooks ist eine verwaltete Umgebung, die auf Jupyter Notebook basiert. Es unterstützt Spark-Magic-Kernel für PySpark, Spark SQL, Spark R und Scala. EMR Notebooks können mit Clustern verwendet werden, die mit Amazon-EMR-Version 5.18.0 und höher erstellt wurden. Weitere Informationen finden Sie unter [Nutzung von EMR Notebooks](#) im Verwaltungshandbuch für Amazon EMR.
- Der S3-optimierte EMRFS-Committer ist beim Schreiben von Parquet-Dateien mit Spark und EMRFS verfügbar. Dieser Committer verbessert die Schreibleistung. Weitere Informationen finden Sie unter [EMRFS-S3-optimierte Committer verwenden](#).

## Änderungen, Verbesserungen und behobene Probleme

- YARN
  - Die Logik, die den Anwendungsmasterprozess auf die Ausführung auf Core-Knoten beschränkt, wurde geändert. Diese Funktionalität verwendet jetzt das Feature und Eigenschaften von YARN-Knotenbeschriftungen in den Klassifizierungen `yarn-site` und der `capacity-scheduler` Konfiguration. Weitere Informationen finden Sie unter <https://docs.aws.amazon.com/emr/latest/ManagementGuide/emr-plan-instances-guidelines.html#emr-plan-spot-YARN>.
- Standard-Amazon-Linux-AMI für Amazon EMR
  - `ruby18`, `php56` und `gcc48` sind nicht mehr standardmäßig installiert. Diese können auf Wunsch mit `yum` installiert werden.
  - Das `aws-sdk` Ruby Gem ist standardmäßig nicht mehr installiert. Es kann, falls gewünscht, mit `gem install aws-sdk` installiert werden. Bestimmte Komponenten können ebenfalls installiert werden. z. B. `gem install aws-sdk-s3`.



## Bekannte Probleme

- EMR Notebooks – Unter bestimmten Umständen kann es vorkommen, dass der Notebook-Editor keine Verbindung zum Cluster herstellen kann, wenn mehrere Notebook-Editoren geöffnet sind. Löschen Sie in diesem Fall die Browser-Cookies und öffnen Sie die Notebook-Editoren erneut.
- CloudWatch ContainerPending Metrische und automatische Skalierung — (In 5.20.0 behoben) Amazon EMR gibt möglicherweise einen negativen Wert für `ContainerPending` Wenn `ContainerPending` in einer Auto-Scaling-Regel verwendet wird, verhält sich das Auto Scaling nicht wie erwartet. Vermeiden Sie die Verwendung von `ContainerPending` mit automatischer Skalierung.
- In den Amazon-EMR-Versionen 5.19.0, 5.20.0 und 5.21.0 werden YARN-Knotenbezeichnungen in einem HDFS-Verzeichnis gespeichert. In einigen Situationen führt dies zu Verzögerungen beim Startup des Core-Knotens und dann zu einem Cluster-Timeout und einem Startfehler. Ab Amazon EMR 5.22.0 ist dieses Problem behoben. YARN-Knotenbezeichnungen werden auf der lokalen Festplatte jedes Clusterknotens gespeichert, wodurch Abhängigkeiten von HDFS vermieden werden.

## 5.19.0 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.2.0	Amazon SageMaker Spark-SDK

Komponente	Version	Beschreibung
emr-ddb	4.7.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.5.1	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.10.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	1.1.0	EMR S3Select-Konnektor
emrfs	2.28.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.6.1	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.

Komponente	Version	Beschreibung
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.5-amzn-0	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.5-amzn-0	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.5-amzn-0	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.8.5-amzn-0	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httpfs-server	2.8.5-amzn-0	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.8.5-amzn-0	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.8.5-amzn-0	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.8.5-amzn-0	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.

Komponente	Version	Beschreibung
hadoop-yarn-resourcemanager	2.8.5-amzn-0	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.5-amzn-0	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.4.7	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.4.7	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.4.7	HBase-Befehlszeilen-Client.
hbase-rest-server	1.4.7	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.4.7	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.3-amzn-2	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.3-amzn-2	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).

Komponente	Version	Beschreibung
hcatalog-webhcat-server	2.3.3-amzn-2	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.3.3-amzn-2	Hive-Befehlszeilen-Client.
hive-hbase	2.3.3-amzn-2	Hive-hbase client.
hive-metastore-server	2.3.3-amzn-2	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.3.3-amzn-2	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.2.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	0.9.4	Multi-User-Server für Jupyter-Notebooks
livy-server	0.5.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP- und Reverse-Proxy-Server.
mahout-client	0.13.0	Bibliothek für Machine Learning.

Komponente	Version	Beschreibung
mxnet	1.3.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.0.0	Oozie-Befehlszeilen-Client.
oozie-server	5.0.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.0-HBase-1.4	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.14.0-HBase-1.4	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.212	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.212	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.

Komponente	Version	Beschreibung
r	3.4.1	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
spark-client	2.3.2	Spark-Befehlszeilen-Clients.
spark-history-server	2.3.2	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.3.2	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.3.2	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	1.11.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.8.0	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.

Komponente	Version	Beschreibung
zookeeper-server	3.4.13	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.13	ZooKeeper Befehlszeilen-Cli-ent.

## 5.19.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

### emr-5.19.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-log4j	Ändert die Werte in der <code>container-log4j.properties</code> -Datei in Hadoop YARN.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.



Klassifizierungen	Beschreibung
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4 j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.

Klassifizierungen	Beschreibung
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.

Klassifizierungen	Beschreibung
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.

Klassifizierungen	Beschreibung
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.

Klassifizierungen	Beschreibung
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-memory	Ändert die Werte in der memory.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
presto-connector-tpcds	Ändert die Werte in der tpcds.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.

Klassifizierungen	Beschreibung
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.18.1

### 5.18.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

### Informationen zur Anwendungsversion

	emr-5.18.1	emr-5.18.0	emr-5.17.2	emr-5.17.1
AWS SDK for Java	1.11.393	1,11.393	1.11.336	1.11.336
Python	2,7, 3,4	2,7, 3,4	2,7, 3,4	2,7, 3,4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.6.0	1.6.0	1.5.2	1.5.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.7	1.4.7	1.4.6	1.4.6
HCatalog	2.3.3	2.3.3	2.3.3	2.3.3
Hadoop	2.8.4	2.8.4	2.8.4	2.8.4
Hive	2.3.3	2.3.3	2.3.3	2.3.3
Hudi	-	-	-	-
Hue	4.2.0	4.2.0	4.2.0	4.2.0
Iceberg	-	-	-	-



	emr-5.18.1	emr-5.18.0	emr-5.17.2	emr-5.17.1
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.8.1	0.8.1	0.8.1	0.8.1
Livy	0.5.0	0.5.0	0.5.0	0.5.0
MXNet	1.2.0	1.2.0	1.2.0	1.2.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.0	4.14.0	4.14.0	4.14.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.210	0.210	0.206	0.206
Spark	2.3.2	2.3.2	2.3.1	2.3.1
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.9.0	1.9.0	1.9.0	1.9.0
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.0	0.8.0	0.7.3	0.7.3
ZooKeeper	3.4.12	3.4.12	3.4.12	3.4.12

## 5.18.1 Versionshinweise

Dies ist eine Patch-Version, um die AWS Signature Version 4-Authentifizierung für Anfragen an Amazon S3 hinzuzufügen. Alle Anwendungen und Komponenten sind dieselben wie in der vorherigen Amazon-EMR-Version.

### ⚠ Important

In dieser Release-Version verwendet Amazon EMR AWS Signature Version 4 ausschließlich zur Authentifizierung von Anfragen an Amazon S3. Weitere Informationen finden Sie unter [Was ist neu](#).

## 5.18.1 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.1.3	Amazon SageMaker Spark-SDK
<code>emr-ddb</code>	4.6.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.5.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-s3-dist-cp	2.10.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	1.1.0	EMR S3Select-Konnektor
emrfs	2.27.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.6.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.4-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.4-amzn-1	HDFS-Service auf Knotenebene zum Speichern von Blöcken.

Komponente	Version	Beschreibung
hadoop-hdfs-library	2.8.4-amzn-1	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.8.4-amzn-1	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httfs-server	2.8.4-amzn-1	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.8.4-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.8.4-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.8.4-amzn-1	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.8.4-amzn-1	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.4-amzn-1	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.

Komponente	Version	Beschreibung
hbase-hmaster	1.4.7	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.4.7	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.4.7	HBase-Befehlszeilen-Client.
hbase-rest-server	1.4.7	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.4.7	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.3-amzn-2	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.3-amzn-2	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.3.3-amzn-2	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.3.3-amzn-2	Hive-Befehlszeilen-Client.
hive-hbase	2.3.3-amzn-2	Hive-hbase client.

Komponente	Version	Beschreibung
hive-metastore-server	2.3.3-amzn-2	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.3.3-amzn-2	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.2.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	0.8.1	Multi-User-Server für Jupyter-Notebooks
livy-server	0.5.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP- und Reverse-Proxy-Server.
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.2.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Toolkit
oozie-client	5.0.0	Oozie-Befehlszeilen-Client.

Komponente	Version	Beschreibung
oozie-server	5.0.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.0-HBase-1.4	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.14.0-HBase-1.4	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.210	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.210	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.1	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
spark-client	2.3.2	Spark-Befehlszeilen-Clients.

Komponente	Version	Beschreibung
spark-history-server	2.3.2	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.3.2	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.3.2	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	1.9.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.8.0	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.12	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.



Komponente	Version	Beschreibung
zookeeper-client	3.4.12	ZooKeeper Befehlszeilen-Client.

### 5.18.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

#### emr-5.18.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-log4j	Ändert die Werte in der <code>container-log4j.properties</code> -Datei in Hadoop YARN.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Ändert die <code>log4j-cli.properties</code> -Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.

Klassifizierungen	Beschreibung
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site

Klassifizierungen	Beschreibung
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.

Klassifizierungen	Beschreibung
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.

Klassifizierungen	Beschreibung
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.

Klassifizierungen	Beschreibung
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.

Klassifizierungen	Beschreibung
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.

Klassifizierungen	Beschreibung
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.18.0

### 5.18.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

### Informationen zur Anwendungsversion

	emr-5.18.0	emr-5.17.2	emr-5.17.1	emr-5.17.0
AWS SDK for Java	1.11.393	1.11.336	1.11.336	1.11.336
Python	2,7, 3,4	2,7, 3,4	2,7, 3,4	2,7, 3,4
Scala	2.11.8	2.11.8	2.11.8	2.11.8



	emr-5.18.0	emr-5.17.2	emr-5.17.1	emr-5.17.0
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.6.0	1.5.2	1.5.2	1.5.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.7	1.4.6	1.4.6	1.4.6
HCatalog	2.3.3	2.3.3	2.3.3	2.3.3
Hadoop	2.8.4	2.8.4	2.8.4	2.8.4
Hive	2.3.3	2.3.3	2.3.3	2.3.3
Hudi	-	-	-	-
Hue	4.2.0	4.2.0	4.2.0	4.2.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.8.1	0.8.1	0.8.1	0.8.1
Livy	0.5.0	0.5.0	0.5.0	0.5.0
MXNet	1.2.0	1.2.0	1.2.0	1.2.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.0	4.14.0	4.14.0	4.14.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0

	emr-5.18.0	emr-5.17.2	emr-5.17.1	emr-5.17.0
Presto	0.210	0.206	0.206	0.206
Spark	2.3.2	2.3.1	2.3.1	2.3.1
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.9.0	1.9.0	1.9.0	1.9.0
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.0	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.12	3.4.12	3.4.12	3.4.12

## 5.18.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.18.0. Änderungen beziehen sich auf Version 5.17.0.

Veröffentlichungsdatum: 24. Oktober 2018

### Upgrades

- Flink 1.6.0
- HBase 1.4.7
- Presto 0.210
- Spark 2.3.2
- Zeppelin 0.8.0

### Neue Features

- Ab Amazon EMR 5.18.0 können Sie das Amazon-EMR-Artefakt-Repository verwenden, um Ihren Auftragscode anhand der genauen Versionen von Bibliotheken und Abhängigkeiten zu erstellen,

die mit bestimmten Amazon-EMR-Versionen verfügbar sind. Weitere Informationen finden Sie unter [Überprüfen von Abhängigkeiten mithilfe des Amazon-EMR-Artefakt-Repositorys](#).

## Änderungen, Verbesserungen und behobene Probleme

- Hive
  - Unterstützung für S3 Select hinzugefügt. Weitere Informationen finden Sie unter [Verwendung von S3 Select mit Hive zur Verbesserung der Leistung](#).
- Presto
  - Unterstützung für [S3 Select](#) Pushdown hinzugefügt. Weitere Informationen finden Sie unter [S3-Select-Pushdown mit Presto zur Verbesserung der Leistung verwenden](#).
- Spark
  - Die standardmäßige log4j-Konfiguration für Spark wurde dahingehend geändert, dass Container-Logs für Spark-Streaming-Aufträge stündlich angezeigt werden. Dadurch wird verhindert, dass Logs für Spark-Streaming-Aufträge mit langer Laufzeit gelöscht werden.

### 5.18.0 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.1.3	Amazon SageMaker Spark-SDK

Komponente	Version	Beschreibung
emr-ddb	4.6.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.5.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.10.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	1.1.0	EMR S3Select-Konnektor
emrfs	2.27.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.6.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.

Komponente	Version	Beschreibung
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.4-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.4-amzn-1	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.4-amzn-1	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.8.4-amzn-1	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httpfs-server	2.8.4-amzn-1	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.8.4-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.8.4-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.8.4-amzn-1	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.

Komponente	Version	Beschreibung
hadoop-yarn-resourcemanager	2.8.4-amzn-1	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.4-amzn-1	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.4.7	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.4.7	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.4.7	HBase-Befehlszeilen-Client.
hbase-rest-server	1.4.7	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.4.7	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.3-amzn-2	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.3-amzn-2	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).

Komponente	Version	Beschreibung
hcatalog-webhcat-server	2.3.3-amzn-2	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.3.3-amzn-2	Hive-Befehlszeilen-Client.
hive-hbase	2.3.3-amzn-2	Hive-hbase client.
hive-metastore-server	2.3.3-amzn-2	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.3.3-amzn-2	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.2.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	0.8.1	Multi-User-Server für Jupyter-Notebooks
livy-server	0.5.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
nginx	1.12.1	nginx [engine x] ist ein HTTP- und Reverse-Proxy-Server.
mahout-client	0.13.0	Bibliothek für Machine Learning.

Komponente	Version	Beschreibung
mxnet	1.2.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.0.0	Oozie-Befehlszeilen-Client.
oozie-server	5.0.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.0-HBase-1.4	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.14.0-HBase-1.4	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.210	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.210	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.



Komponente	Version	Beschreibung
r	3.4.1	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
spark-client	2.3.2	Spark-Befehlszeilen-Clients.
spark-history-server	2.3.2	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.3.2	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.3.2	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	1.9.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.8.0	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.

Komponente	Version	Beschreibung
zookeeper-server	3.4.12	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.12	ZooKeeper Befehlszeilen-Cli-ent.

## 5.18.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

### emr-5.18.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-log4j	Ändert die Werte in der <code>container-log4j.properties</code> -Datei in Hadoop YARN.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.

Klassifizierungen	Beschreibung
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.

Klassifizierungen	Beschreibung
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.

Klassifizierungen	Beschreibung
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.

Klassifizierungen	Beschreibung
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.

Klassifizierungen	Beschreibung
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.



Klassifizierungen	Beschreibung
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.17.2

### 5.17.2 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)

- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

## Informationen zur Anwendungsversion

	emr-5.17.2	emr-5.17.1	emr-5.17.0	emr-5.16.1
AWS SDK for Java	1.11.336	1.11.336	1.11.336	1.11.336
Python	2,7, 3,4	2,7, 3,4	2,7, 3,4	2,7, 3,4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.5.2	1.5.2	1.5.2	1.5.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.6	1.4.6	1.4.6	1.4.4
HCatalog	2.3.3	2.3.3	2.3.3	2.3.3
Hadoop	2.8.4	2.8.4	2.8.4	2.8.4
Hive	2.3.3	2.3.3	2.3.3	2.3.3
Hudi	-	-	-	-
Hue	4.2.0	4.2.0	4.2.0	4.2.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.8.1	0.8.1	0.8.1	0.8.1

	emr-5.17.2	emr-5.17.1	emr-5.17.0	emr-5.16.1
Livy	0.5.0	0.5.0	0.5.0	0.5.0
MXNet	1.2.0	1.2.0	1.2.0	1.2.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.0	4.14.0	4.14.0	4.14.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.206	0.206	0.206	0.203
Spark	2.3.1	2.3.1	2.3.1	2.3.1
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.9.0	1.9.0	1.9.0	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.12	3.4.12	3.4.12	3.4.12

## 5.17.2 Versionshinweise

Dies ist eine Patch-Version, um die AWS Signature Version 4-Authentifizierung für Anfragen an Amazon S3 hinzuzufügen. Alle Anwendungen und Komponenten sind dieselben wie in der vorherigen Amazon-EMR-Version.

### ⚠ Important

In dieser Release-Version verwendet Amazon EMR AWS Signature Version 4 ausschließlich zur Authentifizierung von Anfragen an Amazon S3. Weitere Informationen finden Sie unter [Was ist neu](#).

## 5.17.2 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.1.3	Amazon SageMaker Spark-SDK
<code>emr-ddb</code>	4.6.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.5.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-s3-dist-cp	2.10.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	1.0.0	EMR S3Select-Konnector
emrfs	2.26.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.5.2	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.4-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.4-amzn-1	HDFS-Service auf Knotenebene zum Speichern von Blöcken.

Komponente	Version	Beschreibung
hadoop-hdfs-library	2.8.4-amzn-1	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.8.4-amzn-1	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httfs-server	2.8.4-amzn-1	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.8.4-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.8.4-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.8.4-amzn-1	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.8.4-amzn-1	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.4-amzn-1	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.

Komponente	Version	Beschreibung
hbase-hmaster	1.4.6	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.4.6	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.4.6	HBase-Befehlszeilen-Client.
hbase-rest-server	1.4.6	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.4.6	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.3-amzn-1	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.3-amzn-1	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.3.3-amzn-1	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.3.3-amzn-1	Hive-Befehlszeilen-Client.
hive-hbase	2.3.3-amzn-1	Hive-hbase client.

Komponente	Version	Beschreibung
hive-metastore-server	2.3.3-amzn-1	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.3.3-amzn-1	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.2.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	0.8.1	Multi-User-Server für Jupyter-Notebooks
livy-server	0.5.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.2.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Toolkit
oozie-client	5.0.0	Oozie-Befehlszeilen-Client.



Komponente	Version	Beschreibung
oozie-server	5.0.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.0-HBase-1.4	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.14.0-HBase-1.4	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.206	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.206	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.1	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
spark-client	2.3.1	Spark-Befehlszeilen-Clients.

Komponente	Version	Beschreibung
spark-history-server	2.3.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.3.1	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.3.1	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	1.9.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.7.3	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.12	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.

Komponente	Version	Beschreibung
zookeeper-client	3.4.12	ZooKeeper Befehlszeilen-Client.

## 5.17.2 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

### emr-5.17.2-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-log4j	Ändert die Werte in der <code>container-log4j.properties</code> -Datei in Hadoop YARN.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Ändert die <code>log4j-cli.properties</code> -Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.

Klassifizierungen	Beschreibung
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site

Klassifizierungen	Beschreibung
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.

Klassifizierungen	Beschreibung
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.

Klassifizierungen	Beschreibung
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.

Klassifizierungen	Beschreibung
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.



Klassifizierungen	Beschreibung
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.

Klassifizierungen	Beschreibung
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.17.1

### 5.17.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

### Informationen zur Anwendungsversion

	emr-5.17.1	emr-5.17.0	emr-5.16.1	emr-5.16.0
AWS SDK for Java	1.11.336	1.11.336	1.11.336	1.11.336
Python	2,7, 3,4	2,7, 3,4	2,7, 3,4	2,7, 3,4
Scala	2.11.8	2.11.8	2.11.8	2.11.8

	emr-5.17.1	emr-5.17.0	emr-5.16.1	emr-5.16.0
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.5.2	1.5.2	1.5.0	1.5.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.6	1.4.6	1.4.4	1.4.4
HCatalog	2.3.3	2.3.3	2.3.3	2.3.3
Hadoop	2.8.4	2.8.4	2.8.4	2.8.4
Hive	2.3.3	2.3.3	2.3.3	2.3.3
Hudi	-	-	-	-
Hue	4.2.0	4.2.0	4.2.0	4.2.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.8.1	0.8.1	0.8.1	0.8.1
Livy	0.5.0	0.5.0	0.5.0	0.5.0
MXNet	1.2.0	1.2.0	1.2.0	1.2.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.0	4.14.0	4.14.0	4.14.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0

	emr-5.17.1	emr-5.17.0	emr-5.16.1	emr-5.16.0
Presto	0.206	0.206	0.203	0.203
Spark	2.3.1	2.3.1	2.3.1	2.3.1
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.9.0	1.9.0	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.12	3.4.12	3.4.12	3.4.12

## 5.17.1 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.17.1. Änderungen beziehen sich auf Version 5.17.0.

Erste Version: 18. Juli 2019

Änderungen, Verbesserungen und behobene Probleme

- Das standardmäßige Amazon Linux AMI für Amazon EMR wurde aktualisiert, sodass es wichtige Sicherheitsupdates für den Linux-Kernel enthält, darunter das „TCP SACK Denial of Service Issue“ ([AWS-2019-005](#)).

## 5.17.1 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version

sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-amzn-*EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen myapp-component der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als 2.2-amzn-2 aufgeführt.

Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1.1.3	Amazon SageMaker Spark-SDK
emr-ddb	4.6.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.5.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.10.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	1.0.0	EMR S3Select-Konnector
emrfs	2.26.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.5.2	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.

Komponente	Version	Beschreibung
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.4-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.4-amzn-1	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.4-amzn-1	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.8.4-amzn-1	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httpfs-server	2.8.4-amzn-1	HTTP-Endpunkt für HDFS-Operationen.

Komponente	Version	Beschreibung
hadoop-kms-server	2.8.4-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.8.4-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.8.4-amzn-1	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.8.4-amzn-1	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.4-amzn-1	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.4.6	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.4.6	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.4.6	HBase-Befehlszeilen-Client.

Komponente	Version	Beschreibung
hbase-rest-server	1.4.6	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.4.6	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.3-amzn-1	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.3-amzn-1	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.3.3-amzn-1	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.3.3-amzn-1	Hive-Befehlszeilen-Client.
hive-hbase	2.3.3-amzn-1	Hive-hbase client.
hive-metastore-server	2.3.3-amzn-1	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.3.3-amzn-1	Service zur Annahme von Hive-Abfragen als Webanfragen.



Komponente	Version	Beschreibung
hue-server	4.2.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	0.8.1	Multi-User-Server für Jupyter-Notebooks
livy-server	0.5.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.2.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Toolkit
oozie-client	5.0.0	Oozie-Befehlszeilen-Client.
oozie-server	5.0.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.0-HBase-1.4	Die Phoenix-Bibliotheken für den Server und den Client

Komponente	Version	Beschreibung
phoenix-query-server	4.14.0-HBase-1.4	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.206	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.206	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.1	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
spark-client	2.3.1	Spark-Befehlszeilen-Clients.
spark-history-server	2.3.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.3.1	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.3.1	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.

Komponente	Version	Beschreibung
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	1.9.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.7.3	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.12	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.12	ZooKeeper Befehlszeilen-Client.

### 5.17.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

## emr-5.17.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der capacity-scheduler.xml-Datei in Hadoop.
container-log4j	Ändert die Werte in der container-log4j.properties-Datei in Hadoop YARN.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4j-yarn-session.properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.

Klassifizierungen	Beschreibung
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.

Klassifizierungen	Beschreibung
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.

Klassifizierungen	Beschreibung
jupyter-notebook-conf	Ändert die Werte in der <code>jupyter_notebook_config.py</code> -Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei <code>jupyterhub_config.py</code> .
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der <code>config.json</code> -Datei in Sparkmagic.
livy-conf	Ändert die Werte in der <code>livy.conf</code> -Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei <code>mapred-site.xml</code> der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der <code>oozie-log4j.properties</code> -Datei in Oozie.
oozie-site	Ändert die Werte in der <code>oozie-site.xml</code> -Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der <code>hadoop-metrics2-hbase.properties</code> -Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der <code>hbase-site.xml</code> -Datei in Phoenix.

Klassifizierungen	Beschreibung
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.



Klassifizierungen	Beschreibung
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.

Klassifizierungen	Beschreibung
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.17.0

### 5.17.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

#### Informationen zur Anwendungsversion

	emr-5.17.0	emr-5.16.1	emr-5.16.0	emr-5.15.1
AWS SDK for Java	1.11.336	1.11.336	1.11.336	1.11.333
Python	2,7, 3,4	2,7, 3,4	2,7, 3,4	2,7, 3,4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.5.2	1.5.0	1.5.0	1.4.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2

	emr-5.17.0	emr-5.16.1	emr-5.16.0	emr-5.15.1
HBase	1.4.6	1.4.4	1.4.4	1.4.4
HCatalog	2.3.3	2.3.3	2.3.3	2.3.3
Hadoop	2.8.4	2.8.4	2.8.4	2.8.3
Hive	2.3.3	2.3.3	2.3.3	2.3.3
Hudi	-	-	-	-
Hue	4.2.0	4.2.0	4.2.0	4.2.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.8.1	0.8.1	0.8.1	0.8.1
Livy	0.5.0	0.5.0	0.5.0	0,4,0
MXNet	1.2.0	1.2.0	1.2.0	1.1.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.0	4.14.0	4.14.0	4.13.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.206	0.203	0.203	0.194
Spark	2.3.1	2.3.1	2.3.1	2.3.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.9.0	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4

	emr-5.17.0	emr-5.16.1	emr-5.16.0	emr-5.15.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.12	3.4.12	3.4.12	3.4.12

## 5.17.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.17.0. Änderungen beziehen sich auf Version 5.16.0.

Erste Version: 30. August 2018

### Upgrades

- Flink 1.5.2
- HBase 1.4.6
- Presto 0.206

### Neue Features

- Unterstützung für Tensorflow hinzugefügt. Weitere Informationen finden Sie unter [TensorFlow](#).

### Änderungen, Verbesserungen und behobene Probleme

- JupyterHub
  - Unterstützung für Notebook-Persistenz in Amazon S3 hinzugefügt. Weitere Informationen finden Sie unter [Konfigurieren von Persistenz für Notebooks in Amazon S3](#).
- Spark
  - Unterstützung für [S3 Select](#) hinzugefügt. Weitere Informationen finden Sie unter [S3 Select mit Spark zur Verbesserung der Leistung bei Abfragen verwenden](#).
- Die Probleme mit den Cloudwatch-Metriken und des automatischen Skalierungsfeatures in Amazon-EMR-Version 5.14.0, 5.15.0 oder 5.16.0 wurden behoben.

## Bekannte Probleme

- Wenn Sie einen kerberisierten Cluster mit installiertem Livy erstellen, schlägt Livy mit der Fehlermeldung fehl, dass die einfache Authentifizierung nicht aktiviert ist. Ein Neustart des Livy-Servers behebt das Problem. Um das Problem zu umgehen, fügen Sie während der Clustererstellung einen Schritt `sudo restart livy-server` hinzu, der auf dem Primärknoten ausgeführt wird.
- Wenn Sie ein benutzerdefiniertes Amazon-Linux-AMI verwenden, das auf einem Amazon-Linux-AMI mit einem Erstellungsdatum vom 8.11.2018 basiert, kann der Oozie-Server nicht gestartet werden. Wenn Sie Oozie verwenden, erstellen Sie ein benutzerdefiniertes AMI, das auf einer Amazon-Linux-AMI-ID mit einem anderen Erstellungsdatum basiert. Sie können den folgenden AWS CLI Befehl verwenden, um eine Liste der Image-IDs für alle HVM-Amazon-Linux-AMIs mit einer Version 2018.03 zusammen mit dem Veröffentlichungsdatum zurückzugeben, sodass Sie ein geeignetes Amazon Linux-AMI als Basis auswählen können. MyRegion Ersetzen Sie es durch Ihre Regionskennung, z. B. us-west-2.

```
aws ec2 --region MyRegion describe-images --owner amazon --query 'Images[?
Name!=`null`][[?starts_with(Name, `amzn-ami-hvm-2018.03`) == `true`].
[CreationDate,ImageId,Name]' --output text | sort -rk1
```

### 5.17.0 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-*amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1.1.3	Amazon SageMaker Spark-SDK
emr-ddb	4.6.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.5.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.10.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emr-s3-select	1.0.0	EMR S3Select-Konnector
emrfs	2.26.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.5.2	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken

Komponente	Version	Beschreibung
		aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.4-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.4-amzn-1	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.4-amzn-1	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.8.4-amzn-1	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httpfs-server	2.8.4-amzn-1	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.8.4-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.8.4-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.



Komponente	Version	Beschreibung
hadoop-yarn-nodemanager	2.8.4-amzn-1	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.8.4-amzn-1	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.4-amzn-1	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.4.6	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.4.6	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.4.6	HBase-Befehlszeilen-Client.
hbase-rest-server	1.4.6	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.4.6	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.3-amzn-1	Der "hcat"-Befehlszeilen-Client-für das Bearbeiten des hcatalog-Servers.

Komponente	Version	Beschreibung
hcatalog-server	2.3.3-amzn-1	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.3.3-amzn-1	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.3.3-amzn-1	Hive-Befehlszeilen-Client.
hive-hbase	2.3.3-amzn-1	Hive-hbase client.
hive-metastore-server	2.3.3-amzn-1	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.3.3-amzn-1	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.2.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	0.8.1	Multi-User-Server für Jupyter-Notebooks
livy-server	0.5.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
mahout-client	0.13.0	Bibliothek für Machine Learning.

Komponente	Version	Beschreibung
mxnet	1.2.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.0.0	Oozie-Befehlszeilen-Client.
oozie-server	5.0.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.0-HBase-1.4	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.14.0-HBase-1.4	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.206	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.206	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.

Komponente	Version	Beschreibung
r	3.4.1	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
spark-client	2.3.1	Spark-Befehlszeilen-Clients.
spark-history-server	2.3.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.3.1	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.3.1	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tensorflow	1.9.0	TensorFlow Open-Source-Softwarebibliothek für leistungsstarke numerische Berechnungen.
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.7.3	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.

Komponente	Version	Beschreibung
zookeeper-server	3.4.12	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.12	ZooKeeper Befehlszeilen-Cli-ent.

## 5.17.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

### emr-5.17.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-log4j	Ändert die Werte in der <code>container-log4j.properties</code> -Datei in Hadoop YARN.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.

Klassifizierungen	Beschreibung
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4 j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.

Klassifizierungen	Beschreibung
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.

Klassifizierungen	Beschreibung
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-s3-conf	Konfigurieren Sie die S3-Persistenz für Jupyter Notebooks.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.



Klassifizierungen	Beschreibung
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.

Klassifizierungen	Beschreibung
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.

Klassifizierungen	Beschreibung
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.16.1

### 5.16.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)

- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

## Informationen zur Anwendungsversion

	emr-5.16.1	emr-5.16.0	emr-5.15.1	emr-5.15.0
AWS SDK for Java	1.11.336	1.11.336	1.11.333	1,11.333
Python	2,7, 3,4	2,7, 3,4	2,7, 3,4	2,7, 3,4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.5.0	1.5.0	1.4.2	1.4.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.4	1.4.4	1.4.4	1.4.4
HCatalog	2.3.3	2.3.3	2.3.3	2.3.3
Hadoop	2.8.4	2.8.4	2.8.3	2.8.3
Hive	2.3.3	2.3.3	2.3.3	2.3.3
Hudi	-	-	-	-
Hue	4.2.0	4.2.0	4.2.0	4.2.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.8.1	0.8.1	0.8.1	0.8.1

	emr-5.16.1	emr-5.16.0	emr-5.15.1	emr-5.15.0
Livy	0.5.0	0.5.0	0,4,0	0,4,0
MXNet	1.2.0	1.2.0	1.1.0	1.1.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.0	4.14.0	4.13.0	4.13.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.203	0.203	0.194	0.194
Spark	2.3.1	2.3.1	2.3.0	2.3.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.12	3.4.12	3.4.12	3.4.12

## 5.16.1 Versionshinweise

Dies ist eine Patch-Version, um die AWS Signature Version 4-Authentifizierung für Anfragen an Amazon S3 hinzuzufügen. Alle Anwendungen und Komponenten sind dieselben wie in der vorherigen Amazon-EMR-Version.

### ⚠ Important

In dieser Release-Version verwendet Amazon EMR AWS Signature Version 4 ausschließlich zur Authentifizierung von Anfragen an Amazon S3. Weitere Informationen finden Sie unter [Was ist neu](#).

## 5.16.1 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.1.0	Amazon SageMaker Spark-SDK
<code>emr-ddb</code>	4.6.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.4.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-s3-dist-cp	2.10.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.25.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.5.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.4-amzn-0	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.4-amzn-0	HDFS-Service auf Knotenebene zum Speichern von Blöcken.



Komponente	Version	Beschreibung
hadoop-hdfs-library	2.8.4-amzn-0	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.8.4-amzn-0	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httpfs-server	2.8.4-amzn-0	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.8.4-amzn-0	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.8.4-amzn-0	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.8.4-amzn-0	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.8.4-amzn-0	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.4-amzn-0	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.

Komponente	Version	Beschreibung
hbase-hmaster	1.4.4	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.4.4	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.4.4	HBase-Befehlszeilen-Client.
hbase-rest-server	1.4.4	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.4.4	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.3-amzn-1	Der "hcat"-Befehlszeilen-Client-für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.3-amzn-1	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.3.3-amzn-1	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.3.3-amzn-1	Hive-Befehlszeilen-Client.
hive-hbase	2.3.3-amzn-1	Hive-hbase client.

Komponente	Version	Beschreibung
hive-metastore-server	2.3.3-amzn-1	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.3.3-amzn-1	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.2.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	0.8.1	Multi-User-Server für Jupyter-Notebooks
livy-server	0.5.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.2.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Toolkit
oozie-client	5.0.0	Oozie-Befehlszeilen-Client.

Komponente	Version	Beschreibung
oozie-server	5.0.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.0-HBase-1.4	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.14.0-HBase-1.4	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.203	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.203	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.1	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
spark-client	2.3.1	Spark-Befehlszeilen-Clients.

Komponente	Version	Beschreibung
spark-history-server	2.3.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.3.1	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.3.1	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.7.3	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.12	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.12	ZooKeeper Befehlszeilenclient.

## 5.16.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

### emr-5.16.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-log4j	Ändert die Werte in der <code>container-log4j.properties</code> -Datei in Hadoop YARN.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Ändert die <code>log4j-cli.properties</code> -Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.

Klassifizierungen	Beschreibung
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.

Klassifizierungen	Beschreibung
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.



Klassifizierungen	Beschreibung
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.

Klassifizierungen	Beschreibung
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.

Klassifizierungen	Beschreibung
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.16.0

### 5.16.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

#### Informationen zur Anwendungsversion

	emr-5.16.0	emr-5.15.1	emr-5.15.0	emr-5.14.2
AWS SDK for Java	1.11.336	1.11.333	1,11.333	1.11.297
Python	2,7, 3,4	2,7, 3,4	2,7, 3,4	2,7, 3,4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.5.0	1.4.2	1.4.2	1.4.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2

	emr-5.16.0	emr-5.15.1	emr-5.15.0	emr-5.14.2
HBase	1.4.4	1.4.4	1.4.4	1.4.2
HCatalog	2.3.3	2.3.3	2.3.3	2.3.2
Hadoop	2.8.4	2.8.3	2.8.3	2.8.3
Hive	2.3.3	2.3.3	2.3.3	2.3.2
Hudi	-	-	-	-
Hue	4.2.0	4.2.0	4.2.0	4.1.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.8.1	0.8.1	0.8.1	0.8.1
Livy	0.5.0	0,4,0	0,4,0	0,4,0
MXNet	1.2.0	1.1.0	1.1.0	1.1.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	4.3.0
Phoenix	4.14.0	4.13.0	4.13.0	4.13.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.203	0.194	0.194	0.194
Spark	2.3.1	2.3.0	2.3.0	2.3.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4

	emr-5.16.0	emr-5.15.1	emr-5.15.0	emr-5.14.2
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.12	3.4.12	3.4.12	3.4.10

## 5.16.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.16.0. Änderungen beziehen sich auf Version 5.15.0.

Erste Version: 19. Juli 2018

### Upgrades

- Hadoop 2.8.4
- Flink 1.5.0
- Livy 0.5.0
- MXNet 1.2.0
- Phoenix 4.14.0
- Presto 0.203
- Spark 2.3.1
- AWS SDK for Java 1,11.336
- CUDA 9.2
- Redshift-JDBC-Treiber 1.2.15.1025

### Änderungen, Verbesserungen und behobene Probleme

- HBase
  - [HBASE-20723](#) wurde rückportiert
- Presto
  - Konfigurationsänderungen zur Unterstützung der LDAP-Authentifizierung. Weitere Informationen finden Sie unter [LDAP-Authentifizierung für Presto in Amazon EMR verwenden](#).

- Spark
  - Apache-Spark-Version 2.3.1, verfügbar ab Amazon-EMR-Version 5.16.0, adressiert [CVE-2018-8024](#) und [CVE-2018-1334](#). Wir empfehlen, dass Sie frühere Versionen von Spark zu Spark-Version 2.3.1 oder höher migrieren.

## Bekannte Probleme

- Diese Release-Version unterstützt die Instance-Typen c1.medium oder m1.small nicht. Cluster, die einen dieser Instance-Typen verwenden, können nicht gestartet werden. Um das Problem zu umgehen, geben Sie einen anderen Instance-Typ an oder verwenden Sie eine andere Release-Version.
- Wenn Sie einen kerberisierten Cluster mit installiertem Livy erstellen, schlägt Livy mit der Fehlermeldung fehl, dass die einfache Authentifizierung nicht aktiviert ist. Ein Neustart des Livy-Servers behebt das Problem. Um das Problem zu umgehen, fügen Sie während der Clustererstellung einen Schritt `sudo restart livy-server` hinzu, der auf dem Primärknoten ausgeführt wird.
- Nach dem Neustart des Primärknotens oder des Instance-Controllers werden die CloudWatch Metriken nicht erfasst und die automatische Skalierungsfunktion ist in Amazon EMR Version 5.14.0, 5.15.0 oder 5.16.0 nicht verfügbar. Dieses Problem wurde in Amazon EMR 5.17.0 behoben.

## 5.16.0 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.



Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1.1.0	Amazon SageMaker Spark-SDK
emr-ddb	4.6.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.4.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.10.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.25.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.5.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.

Komponente	Version	Beschreibung
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.4-amzn-0	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.4-amzn-0	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.4-amzn-0	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.8.4-amzn-0	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httpfs-server	2.8.4-amzn-0	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.8.4-amzn-0	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.8.4-amzn-0	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.8.4-amzn-0	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.

Komponente	Version	Beschreibung
hadoop-yarn-resourcemanager	2.8.4-amzn-0	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.4-amzn-0	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.4.4	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.4.4	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.4.4	HBase-Befehlszeilen-Client.
hbase-rest-server	1.4.4	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.4.4	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.3-amzn-1	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.3-amzn-1	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).

Komponente	Version	Beschreibung
hcatalog-webhcat-server	2.3.3-amzn-1	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.3.3-amzn-1	Hive-Befehlszeilen-Client.
hive-hbase	2.3.3-amzn-1	Hive-hbase client.
hive-metastore-server	2.3.3-amzn-1	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.3.3-amzn-1	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.2.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	0.8.1	Multi-User-Server für Jupyter-Notebooks
livy-server	0.5.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.2.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.

Komponente	Version	Beschreibung
nvidia-cuda	9.2.88	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.0.0	Oozie-Befehlszeilen-Client.
oozie-server	5.0.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.14.0-HBase-1.4	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.14.0-HBase-1.4	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.203	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.203	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.1	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
spark-client	2.3.1	Spark-Befehlszeilen-Clients.

Komponente	Version	Beschreibung
spark-history-server	2.3.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.3.1	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.3.1	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.7.3	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.12	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.12	ZooKeeper Befehlszeilenclient.

## 5.16.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

### emr-5.16.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-log4j	Ändert die Werte in der <code>container-log4j.properties</code> -Datei in Hadoop YARN.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Ändert die <code>log4j-cli.properties</code> -Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.

Klassifizierungen	Beschreibung
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.



Klassifizierungen	Beschreibung
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.

Klassifizierungen	Beschreibung
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.

Klassifizierungen	Beschreibung
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-password-authenticator	Ändern Sie Werte in der Presto-Datei password-authenticator.properties.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.

Klassifizierungen	Beschreibung
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.15.1

### 5.15.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

#### Informationen zur Anwendungsversion

	emr-5.15.1	emr-5.15.0	emr-5.14.2	emr-5.14.1
AWS SDK for Java	1.11.333	1,11.333	1.11.297	1,11.297
Python	2,7, 3,4	2,7, 3,4	2,7, 3,4	2,7, 3,4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.4.2	1.4.2	1.4.2	1.4.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2

	emr-5.15.1	emr-5.15.0	emr-5.14.2	emr-5.14.1
HBase	1.4.4	1.4.4	1.4.2	1.4.2
HCatalog	2.3.3	2.3.3	2.3.2	2.3.2
Hadoop	2.8.3	2.8.3	2.8.3	2.8.3
Hive	2.3.3	2.3.3	2.3.2	2.3.2
Hudi	-	-	-	-
Hue	4.2.0	4.2.0	4.1.0	4.1.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.8.1	0.8.1	0.8.1	0.8.1
Livy	0.4.0	0,4,0	0,4,0	0,4,0
MXNet	1.1.0	1.1.0	1.1.0	1.1.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	4.3.0	4.3.0
Phoenix	4.13.0	4.13.0	4.13.0	4.13.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.194	0.194	0.194	0.194
Spark	2.3.0	2.3.0	2.3.0	2.3.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4

	emr-5.15.1	emr-5.15.0	emr-5.14.2	emr-5.14.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.12	3.4.12	3.4.10	3.4.10

### 5.15.1 Versionshinweise

Dies ist eine Patch-Version, um die AWS Signature Version 4-Authentifizierung für Anfragen an Amazon S3 hinzuzufügen. Alle Anwendungen und Komponenten sind dieselben wie in der vorherigen Amazon-EMR-Version.

#### Important

In dieser Release-Version verwendet Amazon EMR AWS Signature Version 4 ausschließlich zur Authentifizierung von Anfragen an Amazon S3. Weitere Informationen finden Sie unter [Was ist neu](#).

### 5.15.1 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit emr oder aws. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen myapp-component der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als 2.2-amzn-2 aufgeführt.



Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1.0.1	Amazon SageMaker Spark-SDK
emr-ddb	4.5.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.4.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.10.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.24.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.4.2	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.

Komponente	Version	Beschreibung
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.3-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.3-amzn-1	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.3-amzn-1	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.8.3-amzn-1	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httpfs-server	2.8.3-amzn-1	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.8.3-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.8.3-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.8.3-amzn-1	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.

Komponente	Version	Beschreibung
hadoop-yarn-resourcemanager	2.8.3-amzn-1	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.3-amzn-1	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.4.4	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.4.4	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.4.4	HBase-Befehlszeilen-Client.
hbase-rest-server	1.4.4	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.4.4	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.3-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.3-amzn-0	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).

Komponente	Version	Beschreibung
hcatalog-webhcat-server	2.3.3-amzn-0	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.3.3-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.3.3-amzn-0	Hive-hbase client.
hive-metastore-server	2.3.3-amzn-0	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.3.3-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.2.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	0.8.1	Multi-User-Server für Jupyter-Notebooks
livy-server	0.4.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.1.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.

Komponente	Version	Beschreibung
nvidia-cuda	9.1.85	Nvidia-Treiber und Cuda-Tool kit
oozie-client	5.0.0	Oozie-Befehlszeilen-Client.
oozie-server	5.0.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.13.0-HBase-1.4	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.13.0-HBase-1.4	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.194	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.194	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.1	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
spark-client	2.3.0	Spark-Befehlszeilen-Clients.

Komponente	Version	Beschreibung
spark-history-server	2.3.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.3.0	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.3.0	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.7.3	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.12	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.12	ZooKeeper Befehlszeilenclient.

## 5.15.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

### emr-5.15.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-log4j	Ändert die Werte in der <code>container-log4j.properties</code> -Datei in Hadoop YARN.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Ändert die <code>log4j-cli.properties</code> -Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.

Klassifizierungen	Beschreibung
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.



Klassifizierungen	Beschreibung
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.

Klassifizierungen	Beschreibung
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.

Klassifizierungen	Beschreibung
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.

Klassifizierungen	Beschreibung
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.15.0

### 5.15.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

#### Informationen zur Anwendungsversion

	emr-5.15.0	emr-5.14.2	emr-5.14.1	emr-5.14.0
AWS SDK for Java	1.11.333	1.11.297	1,11.297	1,11.297
Python	2,7, 3,4	2,7, 3,4	2,7, 3,4	2,7, 3,4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.4.2	1.4.2	1.4.2	1.4.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2

	emr-5.15.0	emr-5.14.2	emr-5.14.1	emr-5.14.0
HBase	1.4.4	1.4.2	1.4.2	1.4.2
HCatalog	2.3.3	2.3.2	2.3.2	2.3.2
Hadoop	2.8.3	2.8.3	2.8.3	2.8.3
Hive	2.3.3	2.3.2	2.3.2	2.3.2
Hudi	-	-	-	-
Hue	4.2.0	4.1.0	4.1.0	4.1.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.8.1	0.8.1	0.8.1	0.8.1
Livy	0.4.0	0,4,0	0,4,0	0,4,0
MXNet	1.1.0	1.1.0	1.1.0	1.1.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	4.3.0	4.3.0	4.3.0
Phoenix	4.13.0	4.13.0	4.13.0	4.13.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.194	0.194	0.194	0.194
Spark	2.3.0	2.3.0	2.3.0	2.3.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4

	emr-5.15.0	emr-5.14.2	emr-5.14.1	emr-5.14.0
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.12	3.4.10	3.4.10	3.4.10

## 5.15.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.15.0. Änderungen beziehen sich auf Version 5.14.0.

Erste Version: 21. Juni 2018

### Upgrades

- Upgrade auf HBase 1.4.4
- Upgrade auf Hive 2.3.3
- Upgrade auf Hue 4.2.0
- Upgrade auf Oozie 5.0.0
- Upgrade auf ZooKeeper 3.4.12.
- AWS SDK auf 1.11.333 aktualisiert

### Änderungen, Verbesserungen und behobene Probleme

- Hive
  - [HIVE-18069](#) wurde rückportiert
- Hue
  - Hue wurde aktualisiert, um sich korrekt bei Livy zu authentifizieren, wenn Kerberos aktiviert ist. Livy wird jetzt unterstützt, wenn Kerberos mit Amazon EMR verwendet wird.
- JupyterHub
  - Aktualisiert, JupyterHub sodass Amazon EMR standardmäßig LDAP-Clientbibliotheken installiert.
  - Ein Fehler im Skript, das selbstsignierte Zertifikate generiert, wurde behoben.



## Bekannte Probleme

- Diese Release-Version unterstützt die Instance-Typen c1.medium oder m1.small nicht. Cluster, die einen dieser Instance-Typen verwenden, können nicht gestartet werden. Um das Problem zu umgehen, geben Sie einen anderen Instance-Typ an oder verwenden Sie eine andere Release-Version.
- Nach dem Neustart des Primärknotens oder des Instance-Controllers werden die CloudWatch Metriken nicht erfasst und die automatische Skalierungsfunktion ist in Amazon EMR Version 5.14.0, 5.15.0 oder 5.16.0 nicht verfügbar. Dieses Problem wurde in Amazon EMR 5.17.0 behoben.

## 5.15.0 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.0.1	Amazon SageMaker Spark-SDK
<code>emr-ddb</code>	4.5.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.4.0	Praktische Bibliotheken für das Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-kinesis	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.10.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.24.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.4.2	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.3-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".

Komponente	Version	Beschreibung
hadoop-hdfs-datanode	2.8.3-amzn-1	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.3-amzn-1	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.8.3-amzn-1	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httpfs-server	2.8.3-amzn-1	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.8.3-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.8.3-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.8.3-amzn-1	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resource-manager	2.8.3-amzn-1	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.

Komponente	Version	Beschreibung
hadoop-yarn-timeline-server	2.8.3-amzn-1	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.4.4	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.4.4	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.4.4	HBase-Befehlszeilen-Client.
hbase-rest-server	1.4.4	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.4.4	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.3-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.3-amzn-0	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.3.3-amzn-0	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.

Komponente	Version	Beschreibung
hive-client	2.3.3-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.3.3-amzn-0	Hive-hbase client.
hive-metastore-server	2.3.3-amzn-0	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.3.3-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.2.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	0.8.1	Multi-User-Server für Jupyter-Notebooks
livy-server	0.4.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.1.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.
nvidia-cuda	9.1.85	Nvidia-Treiber und Cuda-Toolkit

Komponente	Version	Beschreibung
oozie-client	5.0.0	Oozie-Befehlszeilen-Client.
oozie-server	5.0.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.13.0-HBase-1.4	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.13.0-HBase-1.4	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.194	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.194	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.1	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
spark-client	2.3.0	Spark-Befehlszeilen-Clients.

Komponente	Version	Beschreibung
spark-history-server	2.3.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.3.0	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.3.0	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.7.3	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.12	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.12	ZooKeeper Befehlszeilenclient.

## 5.15.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

### emr-5.15.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-log4j	Ändert die Werte in der <code>container-log4j.properties</code> -Datei in Hadoop YARN.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Ändert die <code>log4j-cli.properties</code> -Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.



Klassifizierungen	Beschreibung
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.

Klassifizierungen	Beschreibung
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.

Klassifizierungen	Beschreibung
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.

Klassifizierungen	Beschreibung
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.

Klassifizierungen	Beschreibung
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.14.2

### 5.14.2 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

#### Informationen zur Anwendungsversion

	emr-5.14.2	emr-5.14.1	emr-5.14.0	emr-5.13.1
AWS SDK for Java	1.11.297	1,11.297	1,11.297	1,11.297
Python	2,7, 3,4	2,7, 3,4	2,7, 3,4	2,7, 3,4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.4.2	1.4.2	1.4.2	1.4.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2

	emr-5.14.2	emr-5.14.1	emr-5.14.0	emr-5.13.1
HBase	1.4.2	1.4.2	1.4.2	1.4.2
HCatalog	2.3.2	2.3.2	2.3.2	2.3.2
Hadoop	2.8.3	2.8.3	2.8.3	2.8.3
Hive	2.3.2	2.3.2	2.3.2	2.3.2
Hudi	-	-	-	-
Hue	4.1.0	4.1.0	4.1.0	4.1.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.8.1	0.8.1	0.8.1	-
Livy	0.4.0	0,4,0	0,4,0	0,4,0
MXNet	1.1.0	1.1.0	1.1.0	1.0.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.13.0	4.13.0	4.13.0	4.13.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.194	0.194	0.194	0.194
Spark	2.3.0	2.3.0	2.3.0	2.3.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4



	emr-5.14.2	emr-5.14.1	emr-5.14.0	emr-5.13.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

## 5.14.2 Versionshinweise

Dies ist eine Patch-Version, um die AWS Signature Version 4-Authentifizierung für Anfragen an Amazon S3 hinzuzufügen. Alle Anwendungen und Komponenten sind dieselben wie in der vorherigen Amazon-EMR-Version.

### Important

In dieser Release-Version verwendet Amazon EMR AWS Signature Version 4 ausschließlich zur Authentifizierung von Anfragen an Amazon S3. Weitere Informationen finden Sie unter [Was ist neu](#).

## 5.14.2 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit emr oder aws. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen myapp-component der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als 2.2-amzn-2 aufgeführt.

Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1.0.1	Amazon SageMaker Spark-SDK
emr-ddb	4.5.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.4.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.10.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.23.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.4.2	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.

Komponente	Version	Beschreibung
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.3-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.3-amzn-1	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.3-amzn-1	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.8.3-amzn-1	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httpfs-server	2.8.3-amzn-1	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.8.3-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.8.3-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.8.3-amzn-1	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.

Komponente	Version	Beschreibung
hadoop-yarn-resourcemanager	2.8.3-amzn-1	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.3-amzn-1	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.4.2	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.4.2	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.4.2	HBase-Befehlszeilen-Client.
hbase-rest-server	1.4.2	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.4.2	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.2-amzn-2	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.2-amzn-2	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).

Komponente	Version	Beschreibung
hcatalog-webhcat-server	2.3.2-amzn-2	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.3.2-amzn-2	Hive-Befehlszeilen-Client.
hive-hbase	2.3.2-amzn-2	Hive-hbase client.
hive-metastore-server	2.3.2-amzn-2	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.3.2-amzn-2	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.1.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	0.8.1	Multi-User-Server für Jupyter-Notebooks
livy-server	0.4.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.1.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.

Komponente	Version	Beschreibung
nvidia-cuda	9.1.85	Nvidia-Treiber und Cuda-Tool kit
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.13.0-HBase-1.4	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.13.0-HBase-1.4	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.194	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.194	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.1	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
spark-client	2.3.0	Spark-Befehlszeilen-Clients.

Komponente	Version	Beschreibung
spark-history-server	2.3.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.3.0	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.3.0	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.7.3	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.10	ZooKeeper Befehlszeilenclient.

## 5.14.2 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

### emr-5.14.2-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-log4j	Ändert die Werte in der <code>container-log4j.properties</code> -Datei in Hadoop YARN.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Ändert die <code>log4j-cli.properties</code> -Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.



Klassifizierungen	Beschreibung
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.

Klassifizierungen	Beschreibung
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.

Klassifizierungen	Beschreibung
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.

Klassifizierungen	Beschreibung
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.

Klassifizierungen	Beschreibung
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.14.1

### 5.14.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livvy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

#### Informationen zur Anwendungsversion

	emr-5.14.1	emr-5.14.0	emr-5.13.1	emr-5.13.0
AWS SDK for Java	1.11.297	1,11.297	1,11.297	1,11.297
Python	2,7, 3,4	2,7, 3,4	2,7, 3,4	2,7, 3,4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.4.2	1.4.2	1.4.0	1.4.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2

	emr-5.14.1	emr-5.14.0	emr-5.13.1	emr-5.13.0
HBase	1.4.2	1.4.2	1.4.2	1.4.2
HCatalog	2.3.2	2.3.2	2.3.2	2.3.2
Hadoop	2.8.3	2.8.3	2.8.3	2.8.3
Hive	2.3.2	2.3.2	2.3.2	2.3.2
Hudi	-	-	-	-
Hue	4.1.0	4.1.0	4.1.0	4.1.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.8.1	0.8.1	-	-
Livy	0.4.0	0,4,0	0,4,0	0,4,0
MXNet	1.1.0	1.1.0	1.0.0	1.0.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.13.0	4.13.0	4.13.0	4.13.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.194	0.194	0.194	0.194
Spark	2.3.0	2.3.0	2.3.0	2.3.0
Sqoop	1.4.7	1.4.7	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4



	emr-5.14.1	emr-5.14.0	emr-5.13.1	emr-5.13.0
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

### 5.14.1 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.14.1. Änderungen beziehen sich auf Version 5.14.0.

Veröffentlichungsdatum: 17. Oktober 2018

Das Standard-AMI für Amazon EMR wurde aktualisiert, um potenzielle Sicherheitslücken zu schließen.

#### 5.14.1 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.0.1	Amazon SageMaker Spark-SDK

Komponente	Version	Beschreibung
emr-ddb	4.5.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.4.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.10.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.23.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.4.2	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.

Komponente	Version	Beschreibung
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.3-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.3-amzn-1	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.3-amzn-1	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.8.3-amzn-1	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httpfs-server	2.8.3-amzn-1	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.8.3-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.8.3-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.8.3-amzn-1	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.

Komponente	Version	Beschreibung
hadoop-yarn-resourcemanager	2.8.3-amzn-1	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.3-amzn-1	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.4.2	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.4.2	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.4.2	HBase-Befehlszeilen-Client.
hbase-rest-server	1.4.2	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.4.2	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.2-amzn-2	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.2-amzn-2	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).

Komponente	Version	Beschreibung
hcatalog-webhcat-server	2.3.2-amzn-2	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.3.2-amzn-2	Hive-Befehlszeilen-Client.
hive-hbase	2.3.2-amzn-2	Hive-hbase client.
hive-metastore-server	2.3.2-amzn-2	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.3.2-amzn-2	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.1.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	0.8.1	Multi-User-Server für Jupyter-Notebooks
livy-server	0.4.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.1.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.

Komponente	Version	Beschreibung
nvidia-cuda	9.1.85	Nvidia-Treiber und Cuda-Tool kit
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.13.0-HBase-1.4	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.13.0-HBase-1.4	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.194	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.194	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.1	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
spark-client	2.3.0	Spark-Befehlszeilen-Clients.

Komponente	Version	Beschreibung
spark-history-server	2.3.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.3.0	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.3.0	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.7.3	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.10	ZooKeeper Befehlszeilenclient.

## 5.14.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

### emr-5.14.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-log4j	Ändert die Werte in der <code>container-log4j.properties</code> -Datei in Hadoop YARN.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Ändert die <code>log4j-cli.properties</code> -Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.



Klassifizierungen	Beschreibung
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.

Klassifizierungen	Beschreibung
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.

Klassifizierungen	Beschreibung
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.

Klassifizierungen	Beschreibung
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.

Klassifizierungen	Beschreibung
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.14.0

### 5.14.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

#### Informationen zur Anwendungsversion

	emr-5.14.0	emr-5.13.1	emr-5.13.0	emr-5.12.3
AWS SDK for Java	1.11.297	1,11.297	1,11.297	1.11.267
Python	2,7, 3,4	2,7, 3,4	2,7, 3,4	2,7, 3,4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.4.2	1.4.0	1.4.0	1.4.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2

	emr-5.14.0	emr-5.13.1	emr-5.13.0	emr-5.12.3
HBase	1.4.2	1.4.2	1.4.2	1.4.0
HCatalog	2.3.2	2.3.2	2.3.2	2.3.2
Hadoop	2.8.3	2.8.3	2.8.3	2.8.3
Hive	2.3.2	2.3.2	2.3.2	2.3.2
Hudi	-	-	-	-
Hue	4.1.0	4.1.0	4.1.0	4.1.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.8.1	-	-	-
Livy	0.4.0	0,4,0	0,4,0	0,4,0
MXNet	1.1.0	1.0.0	1.0.0	1.0.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.13.0	4.13.0	4.13.0	4.13.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.194	0.194	0.194	0.188
Spark	2.3.0	2.3.0	2.3.0	2.2.1
Sqoop	1.4.7	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4



	emr-5.14.0	emr-5.13.1	emr-5.13.0	emr-5.12.3
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

## 5.14.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.14.0. Änderungen beziehen sich auf Version 5.13.0.

Erste Version: 4. Juni 2018

### Upgrades

- Upgrade auf Apache Flink 1.4.2
- Apache MXNet auf 1.1.0 aktualisiert
- Apache Sqoop auf 1.4.7 aktualisiert

### Neue Features

- JupyterHub Unterstützung hinzugefügt. Weitere Informationen finden Sie unter [JupyterHub](#).

### Änderungen, Verbesserungen und behobene Probleme

- EMRFS
  - Die userAgent-Zeichenfolge in Anfragen an Amazon S3 wurde aktualisiert und enthält nun die Benutzer- und Gruppeninformationen des aufrufenden Prinzipals. Dies kann zusammen mit AWS CloudTrail Protokollen für eine umfassendere Nachverfolgung von Anfragen verwendet werden.
- HBase
  - [HBASE-20447](#) ist enthalten, das ein Problem behebt, das zu Cache-Problemen führen kann, insbesondere bei geteilten Regionen.
- MXnet

- OpenCV-Plattformversion.
- Spark
  - Wenn Spark Parquet-Dateien mithilfe von EMRFS an einen Amazon S3 S3-Speicherort schreibt, wurde der FileOutputCommitter Algorithmus aktualisiert, sodass er Version 2 statt Version 1 verwendet. Dies reduziert die Anzahl der Umbenennungen, was die Anwendungsleistung verbessert. Diese Änderung beeinträchtigt nicht:
    - Andere Anwendungen als Spark.
    - Anwendungen, die in andere Dateisysteme schreiben, wie HDFS (die immer noch Version 1 von verwenden). FileOutputCommitter
    - Anwendungen, die andere Ausgabeformate wie Text oder CSV verwenden, die bereits EMRFS Direct Write verwenden.

## Bekannte Probleme

- JupyterHub
  - Die Verwendung von Konfigurationsklassifizierungen zur Einrichtung JupyterHub und individuellen Einrichtung von Jupyter-Notebooks beim Erstellen eines Clusters wird nicht unterstützt. Bearbeiten Sie die Dateien `jupyterhub_config.py` und `jupyter_notebook_config.py` für jeden Benutzer manuell. Weitere Informationen finden Sie unter [Konfiguration JupyterHub](#).
  - JupyterHub kann auf Clustern innerhalb eines privaten Subnetzes nicht gestartet werden und schlägt mit der Meldung fehl. `Error: ENOENT: no such file or directory, open '/etc/jupyter/conf/server.crt'` Dies wird durch einen Fehler im Skript verursacht, das selbstsignierte Zertifikate generiert. Verwenden Sie die folgende Problemumgehung zum Generieren selbstsignierter Zertifikate. Alle Befehle werden ausgeführt, während Sie mit dem Primärknoten verbunden sind.

1. Kopieren Sie das Skript zur Zertifikatsgenerierung aus dem Container auf den Primärknoten:

```
sudo docker cp jupyterhub:/tmp/gen_self_signed_cert.sh ./
```

2. Verwenden Sie einen Texteditor, um Zeile 23 zu ändern, um den öffentlichen Hostnamen in den lokalen Hostnamen zu ändern, wie unten gezeigt:

```
local hostname=$(curl -s $EC2_METADATA_SERVICE_URI/local-hostname)
```

3. Führen Sie das Skript aus, um selbstsignierte Zertifikate zu generieren:

```
sudo bash ./gen_self_signed_cert.sh
```

4. Verschieben Sie die Zertifikatsdateien, die das Skript generiert, in das `/etc/jupyter/conf/`-Verzeichnis:

```
sudo mv /tmp/server.crt /tmp/server.key /etc/jupyter/conf/
```

Sie können `tail` die `jupyter.log` Datei überprüfen, um zu überprüfen, ob sie JupyterHub neu gestartet wurde und einen 200-Antwortcode zurückgibt. Beispielsweise:

```
tail -f /var/log/jupyter/jupyter.log
```

Dadurch sollte eine Antwort ähnlich der folgenden ausgegeben:

```
# [I 2018-06-14 18:56:51.356 JupyterHub app:1581] JupyterHub is now running at
https://:9443/
# 19:01:51.359 - info: [ConfigProxy] 200 GET /api/routes
```

- Nach dem Neustart des Primärknotens oder des Instance-Controllers werden die CloudWatch Metriken nicht erfasst und die automatische Skalierungsfunktion ist in Amazon EMR Version 5.14.0, 5.15.0 oder 5.16.0 nicht verfügbar. Dieses Problem wurde in Amazon EMR 5.17.0 behoben.

## 5.14.0 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1.0.1	Amazon SageMaker Spark-SDK
emr-ddb	4.5.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.4.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.10.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.23.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.4.2	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.

Komponente	Version	Beschreibung
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.3-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.3-amzn-1	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.3-amzn-1	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.8.3-amzn-1	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httpfs-server	2.8.3-amzn-1	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.8.3-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.8.3-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.8.3-amzn-1	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.

Komponente	Version	Beschreibung
hadoop-yarn-resourcemanager	2.8.3-amzn-1	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.3-amzn-1	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.4.2	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.4.2	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.4.2	HBase-Befehlszeilen-Client.
hbase-rest-server	1.4.2	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.4.2	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.2-amzn-2	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.2-amzn-2	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).

Komponente	Version	Beschreibung
hcatalog-webhcat-server	2.3.2-amzn-2	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.3.2-amzn-2	Hive-Befehlszeilen-Client.
hive-hbase	2.3.2-amzn-2	Hive-hbase client.
hive-metastore-server	2.3.2-amzn-2	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.3.2-amzn-2	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.1.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
jupyterhub	0.8.1	Multi-User-Server für Jupyter-Notebooks
livy-server	0.4.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.1.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.

Komponente	Version	Beschreibung
nvidia-cuda	9.1.85	Nvidia-Treiber und Cuda-Tool kit
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
opencv	3.4.0	Open Source Computer Vision Library.
phoenix-library	4.13.0-HBase-1.4	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.13.0-HBase-1.4	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.194	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.194	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.1	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
spark-client	2.3.0	Spark-Befehlszeilen-Clients.



Komponente	Version	Beschreibung
spark-history-server	2.3.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.3.0	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.3.0	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.7	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.7.3	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.10	ZooKeeper Befehlszeilenclient.

## 5.14.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

### emr-5.14.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
container-log4j	Ändert die Werte in der <code>container-log4j.properties</code> -Datei in Hadoop YARN.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Ändert die <code>log4j-cli.properties</code> -Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.

Klassifizierungen	Beschreibung
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.

Klassifizierungen	Beschreibung
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.

Klassifizierungen	Beschreibung
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
jupyter-notebook-conf	Ändert die Werte in der jupyter_notebook_config.py-Datei in Jupyter Notebook.
jupyter-hub-conf	Ändern Sie die Werte in JupyterHubs der Datei jupyterhub_config.py.
jupyter-sparkmagic-conf	Ändert die Werte in der config.json-Datei in Sparkmagic.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.

Klassifizierungen	Beschreibung
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.

Klassifizierungen	Beschreibung
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.



## Amazon-EMR-Version 5.13.1

### 5.13.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

#### Informationen zur Anwendungsversion

	emr-5.13.1	emr-5.13.0	emr-5.12.3	emr-5.12.2
AWS SDK for Java	1.11.297	1,11.297	1.11.267	1.11.267
Python	2,7, 3,4	2,7, 3,4	2,7, 3,4	2,7, 3,4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.4.0	1.4.0	1.4.0	1.4.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2

	emr-5.13.1	emr-5.13.0	emr-5.12.3	emr-5.12.2
HBase	1.4.2	1.4.2	1.4.0	1.4.0
HCatalog	2.3.2	2.3.2	2.3.2	2.3.2
Hadoop	2.8.3	2.8.3	2.8.3	2.8.3
Hive	2.3.2	2.3.2	2.3.2	2.3.2
Hudi	-	-	-	-
Hue	4.1.0	4.1.0	4.1.0	4.1.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	0.4.0	0,4,0	0,4,0	0,4,0
MXNet	1.0.0	1.0.0	1.0.0	1.0.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.13.0	4.13.0	4.13.0	4.13.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.194	0.194	0.188	0.188
Spark	2.3.0	2.3.0	2.2.1	2.2.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4

	emr-5.13.1	emr-5.13.0	emr-5.12.3	emr-5.12.2
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

### 5.13.1 Versionshinweise

Dies ist eine Patch-Version, um die AWS Signature Version 4-Authentifizierung für Anfragen an Amazon S3 hinzuzufügen. Alle Anwendungen und Komponenten sind dieselben wie in der vorherigen Amazon-EMR-Version.

#### Important

In dieser Release-Version verwendet Amazon EMR AWS Signature Version 4 ausschließlich zur Authentifizierung von Anfragen an Amazon S3. Weitere Informationen finden Sie unter [Was ist neu](#).

### 5.13.1 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1.0.1	Amazon SageMaker Spark-SDK
emr-ddb	4.5.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.4.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.10.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.22.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.4.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.

Komponente	Version	Beschreibung
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.3-amzn-0	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.3-amzn-0	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.3-amzn-0	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.8.3-amzn-0	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httpfs-server	2.8.3-amzn-0	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.8.3-amzn-0	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.8.3-amzn-0	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.8.3-amzn-0	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.

Komponente	Version	Beschreibung
hadoop-yarn-resourcemanager	2.8.3-amzn-0	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.3-amzn-0	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.4.2	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.4.2	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.4.2	HBase-Befehlszeilen-Client.
hbase-rest-server	1.4.2	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.4.2	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.2-amzn-2	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.2-amzn-2	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).

Komponente	Version	Beschreibung
hcatalog-webhcat-server	2.3.2-amzn-2	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.3.2-amzn-2	Hive-Befehlszeilen-Client.
hive-hbase	2.3.2-amzn-2	Hive-hbase client.
hive-metastore-server	2.3.2-amzn-2	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.3.2-amzn-2	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.1.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
livy-server	0.4.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.0.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.
nvidia-cuda	9.1.85	Nvidia-Treiber und Cuda-Toolkit

Komponente	Version	Beschreibung
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.13.0-HBase-1.4	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.13.0-HBase-1.4	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.194	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.194	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.1	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
spark-client	2.3.0	Spark-Befehlszeilen-Clients.
spark-history-server	2.3.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.



Komponente	Version	Beschreibung
spark-on-yarn	2.3.0	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.3.0	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.7.3	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.10	ZooKeeper Befehlszeilenclient.

### 5.13.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

## emr-5.13.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der capacity-scheduler.xml-Datei in Hadoop.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4 j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.

Klassifizierungen	Beschreibung
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.

Klassifizierungen	Beschreibung
hive-parquet-logging	Ändert die Werte in der <code>parquet-logging.properties</code> -Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von <code>hive-exec-log</code> Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von Hive. <code>llap-daemon-log</code>
hive-log4j2	Ändert die Werte in der <code>hive-log4j2.properties</code> -Datei in Hive.
hive-site	Ändert die Werte in der <code>hive-site.xml</code> -Datei in Hive.
hiveserver2-site	Ändert die Werte in der <code>hiveserver2-site.xml</code> -Datei von <code>Server2</code> in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der <code>httpfs-site.xml</code> -Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der <code>kms-acls.xml</code> -Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der <code>kms-log4j.properties</code> -Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der <code>kms-site.xml</code> -Datei in Hadoop.

Klassifizierungen	Beschreibung
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der Anwendungsumgebung MapReduce .
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.

Klassifizierungen	Beschreibung
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.

Klassifizierungen	Beschreibung
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.13.0

### 5.13.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)



- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

## Informationen zur Anwendungsversion

	emr-5.13.0	emr-5.12.3	emr-5.12.2	emr-5.12.1
AWS SDK for Java	1.11.297	1.11.267	1.11.267	1.11.267
Python	2,7, 3,4	2,7, 3,4	2,7, 3,4	2,7, 3,4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.4.0	1.4.0	1.4.0	1.4.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.2	1.4.0	1.4.0	1.4.0
HCatalog	2.3.2	2.3.2	2.3.2	2.3.2
Hadoop	2.8.3	2.8.3	2.8.3	2.8.3
Hive	2.3.2	2.3.2	2.3.2	2.3.2
Hudi	-	-	-	-
Hue	4.1.0	4.1.0	4.1.0	4.1.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-

	emr-5.13.0	emr-5.12.3	emr-5.12.2	emr-5.12.1
Livy	0.4.0	0,4,0	0,4,0	0,4,0
MXNet	1.0.0	1.0.0	1.0.0	1.0.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.13.0	4.13.0	4.13.0	4.13.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.194	0.188	0.188	0.188
Spark	2.3.0	2.2.1	2.2.1	2.2.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

## 5.13.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.13.0. Änderungen beziehen sich auf Version 5.12.0.

### Upgrades

- Upgrade auf Spark 2.3.0
- Upgrade auf HBase 1.4.2

- Upgrade auf Presto 0.194
- Auf AWS SDK for Java 1.11.297 aktualisiert

## Änderungen, Verbesserungen und behobene Probleme

- Hive
  - [HIVE-15436](#) wurde rückportiert. Verbesserte Hive-APIs, sodass nur Ansichten zurückgegeben werden.

## Bekannte Probleme

- MXNet hat derzeit keine OpenCV-Bibliotheken.

## 5.13.0 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.0.1	Amazon SageMaker Spark-SDK
<code>emr-ddb</code>	4.5.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-goodies	2.4.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.10.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.22.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.4.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.

Komponente	Version	Beschreibung
hadoop-client	2.8.3-amzn-0	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.3-amzn-0	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.3-amzn-0	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.8.3-amzn-0	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httpfs-server	2.8.3-amzn-0	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.8.3-amzn-0	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.8.3-amzn-0	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.8.3-amzn-0	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.8.3-amzn-0	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.

Komponente	Version	Beschreibung
hadoop-yarn-timeline-server	2.8.3-amzn-0	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.4.2	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.4.2	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.4.2	HBase-Befehlszeilen-Client.
hbase-rest-server	1.4.2	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.4.2	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.2-amzn-2	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.2-amzn-2	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.3.2-amzn-2	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.

Komponente	Version	Beschreibung
hive-client	2.3.2-amzn-2	Hive-Befehlszeilen-Client.
hive-hbase	2.3.2-amzn-2	Hive-hbase client.
hive-metastore-server	2.3.2-amzn-2	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.3.2-amzn-2	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.1.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
livy-server	0.4.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.0.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.
nvidia-cuda	9.1.85	Nvidia-Treiber und Cuda-Toolkit
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.

Komponente	Version	Beschreibung
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.13.0-HBase-1.4	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.13.0-HBase-1.4	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.194	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.194	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
r	3.4.1	The R Project for Statistical Computing (Software zur statistischen Datenverarbeitung)
spark-client	2.3.0	Spark-Befehlszeilen-Clients.
spark-history-server	2.3.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.



Komponente	Version	Beschreibung
spark-on-yarn	2.3.0	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.3.0	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.7.3	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.10	ZooKeeper Befehlszeilenclient.

### 5.13.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

## emr-5.13.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der capacity-scheduler.xml-Datei in Hadoop.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4 j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.

Klassifizierungen	Beschreibung
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.

Klassifizierungen	Beschreibung
hive-parquet-logging	Ändert die Werte in der <code>parquet-logging.properties</code> -Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von <code>hive-exec-log</code> Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von Hive. <code>llap-daemon-log</code>
hive-log4j2	Ändert die Werte in der <code>hive-log4j2.properties</code> -Datei in Hive.
hive-site	Ändert die Werte in der <code>hive-site.xml</code> -Datei in Hive.
hiveserver2-site	Ändert die Werte in der <code>hiveserver2-site.xml</code> -Datei von <code>Server2</code> in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der <code>httpfs-site.xml</code> -Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der <code>kms-acls.xml</code> -Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der <code>kms-log4j.properties</code> -Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der <code>kms-site.xml</code> -Datei in Hadoop.

Klassifizierungen	Beschreibung
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der Anwendungsumgebung MapReduce .
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.

Klassifizierungen	Beschreibung
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.

Klassifizierungen	Beschreibung
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.12.3

### 5.12.3 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)



- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

## Informationen zur Anwendungsversion

	emr-5.12.3	emr-5.12.2	emr-5.12.1	emr-5.12.0
AWS SDK for Java	1.11.267	1.11.267	1.11.267	1.11.267
Python	2,7, 3,4	2,7, 3,4	2,7, 3,4	2,7, 3,4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.4.0	1.4.0	1.4.0	1.4.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.0	1.4.0	1.4.0	1.4.0
HCatalog	2.3.2	2.3.2	2.3.2	2.3.2
Hadoop	2.8.3	2.8.3	2.8.3	2.8.3
Hive	2.3.2	2.3.2	2.3.2	2.3.2
Hudi	-	-	-	-
Hue	4.1.0	4.1.0	4.1.0	4.1.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-

	emr-5.12.3	emr-5.12.2	emr-5.12.1	emr-5.12.0
Livy	0.4.0	0,4,0	0,4,0	0,4,0
MXNet	1.0.0	1.0.0	1.0.0	1.0.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.13.0	4.13.0	4.13.0	4.13.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.188	0.188	0.188	0.188
Spark	2.2.1	2.2.1	2.2.1	2.2.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

### 5.12.3 Versionshinweise

Dies ist eine Patch-Version, um die AWS Signature Version 4-Authentifizierung für Anfragen an Amazon S3 hinzuzufügen. Alle Anwendungen und Komponenten sind dieselben wie in der vorherigen Amazon-EMR-Version.

### Important

In dieser Release-Version verwendet Amazon EMR AWS Signature Version 4 ausschließlich zur Authentifizierung von Anfragen an Amazon S3. Weitere Informationen finden Sie unter [Was ist neu](#).

## 5.12.3 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-`amzn-EmrVersion`. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.0.1	Amazon SageMaker Spark-SDK
<code>emr-ddb</code>	4.5.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.4.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-s3-dist-cp	2.9.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.21.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.4.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.3-amzn-0	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.3-amzn-0	HDFS-Service auf Knotenebene zum Speichern von Blöcken.

Komponente	Version	Beschreibung
hadoop-hdfs-library	2.8.3-amzn-0	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.8.3-amzn-0	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httfs-server	2.8.3-amzn-0	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.8.3-amzn-0	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.8.3-amzn-0	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.8.3-amzn-0	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.8.3-amzn-0	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.3-amzn-0	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.

Komponente	Version	Beschreibung
hbase-hmaster	1.4.0	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.4.0	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.4.0	HBase-Befehlszeilen-Client.
hbase-rest-server	1.4.0	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.4.0	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.2-amzn-1	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.2-amzn-1	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.3.2-amzn-1	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.3.2-amzn-1	Hive-Befehlszeilen-Client.
hive-hbase	2.3.2-amzn-1	Hive-hbase client.

Komponente	Version	Beschreibung
hive-metastore-server	2.3.2-amzn-1	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.3.2-amzn-1	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.1.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
livy-server	0.4.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.0.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.
nvidia-cuda	9.1.85	Nvidia-Treiber und Cuda-Toolkit
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.

Komponente	Version	Beschreibung
phoenix-library	4.13.0-HBase-1.4	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.13.0-HBase-1.4	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.188	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.188	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
spark-client	2.2.1	Spark-Befehlszeilen-Clients.
spark-history-server	2.2.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.2.1	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.2.1	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.



Komponente	Version	Beschreibung
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.7.3	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.10	ZooKeeper Befehlszeilenclient.

### 5.12.3 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

#### emr-5.12.3-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.

Klassifizierungen	Beschreibung
flink-conf	Ändert die flink-conf.yaml-Einstellungen.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4 j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.

Klassifizierungen	Beschreibung
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log

Klassifizierungen	Beschreibung
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der Anwendungsumgebung MapReduce .

Klassifizierungen	Beschreibung
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.

Klassifizierungen	Beschreibung
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.12.2

### 5.12.2 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)



## Informationen zur Anwendungsversion

	emr-5.12.2	emr-5.12.1	emr-5.12.0	emr-5.11.4
AWS SDK for Java	1.11.267	1.11.267	1.11.267	1.11.238
Python	2,7, 3,4	2,7, 3,4	2,7, 3,4	2,7, 3,4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.4.0	1.4.0	1.4.0	1.3.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.0	1.4.0	1.4.0	1.3.1
HCatalog	2.3.2	2.3.2	2.3.2	2.3.2
Hadoop	2.8.3	2.8.3	2.8.3	2.7.3
Hive	2.3.2	2.3.2	2.3.2	2.3.2
Hudi	-	-	-	-
Hue	4.1.0	4.1.0	4.1.0	4.0.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	0.4.0	0,4,0	0,4,0	0,4,0
MXNet	1.0.0	1.0.0	1.0.0	0.12.0

	emr-5.12.2	emr-5.12.1	emr-5.12.0	emr-5.11.4
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.13.0	4.13.0	4.13.0	4.11.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.188	0.188	0.188	0.187
Spark	2.2.1	2.2.1	2.2.1	2.2.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

## 5.12.2 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.12.2.

Änderungen beziehen sich auf Version 5.12.1

Erste Version: 29. August 2018

Änderungen, Verbesserungen und behobene Probleme

- Diese Version behebt eine potenzielle Schwachstellen.

## 5.12.2 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.0.1	Amazon SageMaker Spark-SDK
<code>emr-ddb</code>	4.5.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.4.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-s3-dist-cp</code>	2.9.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
<code>emrfs</code>	2.21.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
flink-client	1.4.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.3-amzn-0	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.3-amzn-0	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.3-amzn-0	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.8.3-amzn-0	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.

Komponente	Version	Beschreibung
hadoop-https-server	2.8.3-amzn-0	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.8.3-amzn-0	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.8.3-amzn-0	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.8.3-amzn-0	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resource-manager	2.8.3-amzn-0	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.3-amzn-0	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.4.0	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.4.0	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.

Komponente	Version	Beschreibung
hbase-client	1.4.0	HBase-Befehlszeilen-Client.
hbase-rest-server	1.4.0	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.4.0	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.2-amzn-1	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.2-amzn-1	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.3.2-amzn-1	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.3.2-amzn-1	Hive-Befehlszeilen-Client.
hive-hbase	2.3.2-amzn-1	Hive-hbase client.
hive-metastore-server	2.3.2-amzn-1	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.3.2-amzn-1	Service zur Annahme von Hive-Abfragen als Webanfragen.

Komponente	Version	Beschreibung
hue-server	4.1.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
livy-server	0.4.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.0.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.
nvidia-cuda	9.1.85	Nvidia-Treiber und Cuda-Toolkit
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.13.0-HBase-1.4	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.13.0-HBase-1.4	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.

Komponente	Version	Beschreibung
presto-coordinator	0.188	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.188	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
spark-client	2.2.1	Spark-Befehlszeilen-Clients.
spark-history-server	2.2.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.2.1	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.2.1	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.7.3	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.



Komponente	Version	Beschreibung
zookeeper-server	3.4.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.10	ZooKeeper Befehlszeilenclient.

## 5.12.2 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

### emr-5.12.2-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink <code>log4j-yarn-session.properties</code> .

Klassifizierungen	Beschreibung
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.

Klassifizierungen	Beschreibung
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.

Klassifizierungen	Beschreibung
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der Anwendungsumgebung MapReduce .
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.

Klassifizierungen	Beschreibung
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-cassandra	Ändert die Werte in der <code>cassandra.properties</code> -Datei in Presto.
presto-connector-hive	Ändert die Werte in der <code>hive.properties</code> -Datei in Presto.
presto-connector-jmx	Ändert die Werte in der <code>jmx.properties</code> -Datei in Presto.
presto-connector-kafka	Ändert die Werte in der <code>kafka.properties</code> -Datei in Presto.
presto-connector-localfile	Ändert die Werte in der <code>localfile.properties</code> -Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der <code>mongodb.properties</code> -Datei in Presto.
presto-connector-mysql	Ändert die Werte in der <code>mysql.properties</code> -Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der <code>postgresql.properties</code> -Datei in Presto.
presto-connector-raptor	Ändert die Werte in der <code>raptor.properties</code> -Datei in Presto.
presto-connector-redis	Ändert die Werte in der <code>redis.properties</code> -Datei in Presto.
presto-connector-redshift	Ändert die Werte in der <code>redshift.properties</code> -Datei.
presto-connector-tpch	Ändert die Werte in der <code>tpch.properties</code> -Datei in Presto.

Klassifizierungen	Beschreibung
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.

Klassifizierungen	Beschreibung
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.12.1

### 5.12.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

### Informationen zur Anwendungsversion

	emr-5.12.1	emr-5.12.0	emr-5.11.4	emr-5.11.3
AWS SDK for Java	1.11.267	1.11.267	1.11.238	1,11.238
Python	2,7, 3,4	2,7, 3,4	2,7, 3,4	2,7, 3,4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-



	emr-5.12.1	emr-5.12.0	emr-5.11.4	emr-5.11.3
Delta	-	-	-	-
Flink	1.4.0	1.4.0	1.3.2	1.3.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.0	1.4.0	1.3.1	1.3.1
HCatalog	2.3.2	2.3.2	2.3.2	2.3.2
Hadoop	2.8.3	2.8.3	2.7.3	2.7.3
Hive	2.3.2	2.3.2	2.3.2	2.3.2
Hudi	-	-	-	-
Hue	4.1.0	4.1.0	4.0.1	4.0.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	0.4.0	0,4,0	0,4,0	0,4,0
MXNet	1.0.0	1.0.0	0.12.0	0.12.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.13.0	4.13.0	4.11.0	4.11.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.188	0.188	0.187	0.187
Spark	2.2.1	2.2.1	2.2.1	2.2.1

	emr-5.12.1	emr-5.12.0	emr-5.11.4	emr-5.11.3
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

## 5.12.1 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.12.1. Änderungen beziehen sich auf Version 5.12.0.

Erste Version: 29. März 2018

Änderungen, Verbesserungen und behobene Probleme

- Der Amazon-Linux-Kernel des standardmäßigen Amazon-Linux-AMI für Amazon EMR wurde aktualisiert, um potenzielle Schwachstellen zu schließen.

## 5.12.1 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn zum Beispiel eine Open-Source-

Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1.0.1	Amazon SageMaker Spark-SDK
<code>emr-ddb</code>	4.5.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.4.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-s3-dist-cp</code>	2.9.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
<code>emrfs</code>	2.21.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>flink-client</code>	1.4.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
<code>ganglia-monitor</code>	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.

Komponente	Version	Beschreibung
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.3-amzn-0	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.3-amzn-0	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.3-amzn-0	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.8.3-amzn-0	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httpfs-server	2.8.3-amzn-0	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.8.3-amzn-0	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider

Komponente	Version	Beschreibung
hadoop-mapred	2.8.3-amzn-0	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.8.3-amzn-0	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.8.3-amzn-0	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.3-amzn-0	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.4.0	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.4.0	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.4.0	HBase-Befehlszeilen-Client.
hbase-rest-server	1.4.0	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.4.0	Service, der einen Thrift-Endpunkt für HBase bereitstellt.

Komponente	Version	Beschreibung
hcatalog-client	2.3.2-amzn-1	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.2-amzn-1	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.3.2-amzn-1	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.3.2-amzn-1	Hive-Befehlszeilen-Client.
hive-hbase	2.3.2-amzn-1	Hive-hbase client.
hive-metastore-server	2.3.2-amzn-1	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.3.2-amzn-1	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.1.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
livy-server	0.4.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
mahout-client	0.13.0	Bibliothek für Machine Learning.

Komponente	Version	Beschreibung
mxnet	1.0.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.
nvidia-cuda	9.1.85	Nvidia-Treiber und Cuda-Tool kit
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.13.0-HBase-1.4	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.13.0-HBase-1.4	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.188	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.188	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
spark-client	2.2.1	Spark-Befehlszeilen-Clients.

Komponente	Version	Beschreibung
spark-history-server	2.2.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.2.1	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.2.1	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.7.3	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.10	ZooKeeper Befehlszeilenclient.



## 5.12.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

### emr-5.12.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Ändert die <code>log4j-cli.properties</code> -Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.

Klassifizierungen	Beschreibung
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.

Klassifizierungen	Beschreibung
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.

Klassifizierungen	Beschreibung
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der Anwendungsumgebung MapReduce .
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.

Klassifizierungen	Beschreibung
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.

Klassifizierungen	Beschreibung
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.12.0

### 5.12.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

### Informationen zur Anwendungsversion

	emr-5.12.0	emr-5.11.4	emr-5.11.3	emr-5.11.2
AWS SDK for Java	1.11.267	1.11.238	1,11.238	1,11.238
Python	2,7, 3,4	2,7, 3,4	2,7, 3,4	2,7, 3,4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.4.0	1.3.2	1.3.2	1.3.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.0	1.3.1	1.3.1	1.3.1
HCatalog	2.3.2	2.3.2	2.3.2	2.3.2
Hadoop	2.8.3	2.7.3	2.7.3	2.7.3
Hive	2.3.2	2.3.2	2.3.2	2.3.2
Hudi	-	-	-	-
Hue	4.1.0	4.0.1	4.0.1	4.0.1
Iceberg	-	-	-	-



	emr-5.12.0	emr-5.11.4	emr-5.11.3	emr-5.11.2
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	0.4.0	0,4,0	0,4,0	0,4,0
MXNet	1.0.0	0.12.0	0.12.0	0.12.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.13.0	4.11.0	4.11.0	4.11.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.188	0.187	0.187	0.187
Spark	2.2.1	2.2.1	2.2.1	2.2.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

## 5.12.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.12.0. Änderungen beziehen sich auf Version 5.11.1.

## Upgrades

- AWS SDK for Java 1.11.238 ⇒ 1.11.267. Weitere Informationen finden Sie im [AWS SDK for Java Change Log](#) on GitHub.
- Hadoop 2.7.3 ⇒ 2.8.3. Weitere Informationen finden Sie unter [Apache Hadoop im Releases](#).
- Flink 1.3.2 ⇒ 1.4.0. Weitere Informationen finden Sie in der [Versionsankündigung zu Apache Flink 1.4.0](#).
- HBase 1.3.1 ⇒ 1.4.0. Weitere Informationen finden Sie in der [HBase Versionsankündigung](#)..
- Hue 4.0.1 ⇒ 4.1.0. Weitere Informationen finden Sie in den [Versionshinweisen](#).
- MxNet 0.12.0 ⇒ 1.0.0. Weitere Informationen finden Sie im [MXNet Change Log](#) on GitHub.
- Presto 0.187 ⇒ 0.188. Weitere Informationen finden Sie in den [Versionshinweisen](#).

## Änderungen, Verbesserungen und behobene Probleme

- Hadoop
  - Die `yarn.resourcemanager.decommissioning.timeout`-Eigenschaft wurde zu `yarn.resourcemanager.nodemanager-graceful-decommission-timeout-secs` geändert. Sie können diese Eigenschaft verwenden, um die Cluster-Skalierung anzupassen. Weitere Informationen finden Sie unter [Cluster Sale down](#) im Verwaltungshandbuch für Amazon EMR.
  - Die Hadoop-CLI hat die `-d`-Option zum Befehl `cp` (kopieren) hinzugefügt, der das direkte Kopieren spezifiziert. Sie können dies verwenden, um die Erstellung einer `.COPYING`-Zwischendatei zu vermeiden, wodurch das Kopieren von Daten zwischen Amazon S3 beschleunigt wird. Weitere Informationen finden Sie unter [HADOOP-12384](#).
- Pig
  - Die `pig-env`-Konfigurationsklassifizierung wurde hinzugefügt, die die Konfiguration der Pig-Umgebungseigenschaften vereinfacht. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).
- Presto
  - Die Konfigurationsklassifizierung `presto-connector-redshift` wurde hinzugefügt, mit der Sie Werte in der Presto `redshift.properties` konfigurieren können. Weitere Informationen finden Sie unter [Redshift Connector](#) in der Presto-Dokumentation und [Anwendungen konfigurieren](#).

- Presto-Unterstützung für EMRFS wurde hinzugefügt und ist die Standardkonfiguration. Frühere Amazon EMR-Versionen verwendeten PrestoS3FileSystem, was die einzige Option war. Weitere Informationen finden Sie unter [EMRFS- und PrestoS3-Konfiguration FileSystem](#).

#### Note

Wenn Sie die zugrunde liegenden Daten in Amazon S3 mit Amazon-EMR-Version 5.12.0 abfragen, können Presto-Fehler auftreten. Dies liegt daran, dass Presto die Konfigurationsklassifizierungswerte von `emrfs-site.xml` nicht abrufen. Um dieses Problem zu umgehen, erstellen Sie ein `emrfs`-Unterverzeichnis unter `usr/lib/presto/plugin/hive-hadoop2/` und erstellen einen Symlink `usr/lib/presto/plugin/hive-hadoop2/emrfs` zu der vorhandenen Datei `/usr/share/aws/emr/emrfs/conf/emrfs-site.xml`. Starten Sie dann den Presto-Server-Prozess neu (`sudo presto-server stop` gefolgt von `sudo presto-server start`).

- Spark
  - [SPARK-22036](#) rückportiert: Multiplikation gibt manchmal Null zurück. BigDecimal

#### Bekanntere Probleme

- MXNet enthält keine OpenCV-Bibliotheken.
- SparkR ist nicht für Cluster verfügbar, die mit einem benutzerdefinierten AMI erstellt wurden, da R nicht standardmäßig auf Clusterknoten installiert ist.

## 5.12.0 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die

Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als 2.2-amzn-2 aufgeführt.

Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1.0.1	Amazon SageMaker Spark-SDK
emr-ddb	4.5.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.4.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.9.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.21.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.4.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken

Komponente	Version	Beschreibung
		aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.8.3-amzn-0	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.8.3-amzn-0	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.8.3-amzn-0	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.8.3-amzn-0	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httpfs-server	2.8.3-amzn-0	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.8.3-amzn-0	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.8.3-amzn-0	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.

Komponente	Version	Beschreibung
hadoop-yarn-nodemanager	2.8.3-amzn-0	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.8.3-amzn-0	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.8.3-amzn-0	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.4.0	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.4.0	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.4.0	HBase-Befehlszeilen-Client.
hbase-rest-server	1.4.0	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.4.0	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.2-amzn-1	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.

Komponente	Version	Beschreibung
hcatalog-server	2.3.2-amzn-1	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.3.2-amzn-1	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.3.2-amzn-1	Hive-Befehlszeilen-Client.
hive-hbase	2.3.2-amzn-1	Hive-hbase client.
hive-metastore-server	2.3.2-amzn-1	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.3.2-amzn-1	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.1.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
livy-server	0.4.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	1.0.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.

Komponente	Version	Beschreibung
mysql-server	5.5.54+	MySQL-Datenbankserver.
nvidia-cuda	9.1.85	Nvidia-Treiber und Cuda-Tool kit
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.13.0-HBase-1.4	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.13.0-HBase-1.4	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.188	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.188	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
spark-client	2.2.1	Spark-Befehlszeilen-Clients.
spark-history-server	2.2.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.



Komponente	Version	Beschreibung
spark-on-yarn	2.2.1	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.2.1	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.7.3	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.10	ZooKeeper Befehlszeilenclient.

## 5.12.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

## emr-5.12.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der capacity-scheduler.xml-Datei in Hadoop.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4 j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.

Klassifizierungen	Beschreibung
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.

Klassifizierungen	Beschreibung
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.

Klassifizierungen	Beschreibung
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der Anwendungsumgebung MapReduce .
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-env	Ändert die Werte in der Pig-Umgebung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.

Klassifizierungen	Beschreibung
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-redshift	Ändert die Werte in der redshift.properties-Datei.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.

Klassifizierungen	Beschreibung
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.11.4

### 5.11.4 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)



- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

## Informationen zur Anwendungsversion

	emr-5.11.4	emr-5.11.3	emr-5.11.2	emr-5.11.1
AWS SDK for Java	1.11.238	1,11.238	1,11.238	1,11.238
Python	2,7, 3,4	2,7, 3,4	2,7, 3,4	2,7, 3,4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.3.2	1.3.2	1.3.2	1.3.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.1	1.3.1	1.3.1
HCatalog	2.3.2	2.3.2	2.3.2	2.3.2
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.3.2	2.3.2	2.3.2	2.3.2
Hudi	-	-	-	-
Hue	4.0.1	4.0.1	4.0.1	4.0.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-

	emr-5.11.4	emr-5.11.3	emr-5.11.2	emr-5.11.1
Livy	0.4.0	0,4,0	0,4,0	0,4,0
MXNet	0.12.0	0.12.0	0.12.0	0.12.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.11.0	4.11.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.187	0.187	0.187	0.187
Spark	2.2.1	2.2.1	2.2.1	2.2.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

## 5.11.4 Versionshinweise

Dies ist eine Patch-Version, um die AWS Signature Version 4-Authentifizierung für Anfragen an Amazon S3 hinzuzufügen. Alle Anwendungen und Komponenten sind dieselben wie in der vorherigen Amazon-EMR-Version.

### ⚠ Important

In dieser Release-Version verwendet Amazon EMR AWS Signature Version 4 ausschließlich zur Authentifizierung von Anfragen an Amazon S3. Weitere Informationen finden Sie unter [Was ist neu](#).

## 5.11.4 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1,0	Amazon SageMaker Spark-SDK
<code>emr-ddb</code>	4.5.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.4.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-s3-dist-cp	2.8.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.20.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.3.2	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-6	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-6	HDFS-Service auf Knotenebene zum Speichern von Blöcken.

Komponente	Version	Beschreibung
hadoop-hdfs-library	2.7.3-amzn-6	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.7.3-amzn-6	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httfs-server	2.7.3-amzn-6	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.7.3-amzn-6	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.7.3-amzn-6	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-6	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.3-amzn-6	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-6	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.

Komponente	Version	Beschreibung
hbase-hmaster	1.3.1	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.3.1	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.3.1	HBase-Befehlszeilen-Client.
hbase-rest-server	1.3.1	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.3.1	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.2-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.2-amzn-0	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.3.2-amzn-0	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.3.2-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.3.2-amzn-0	Hive-hbase client.

Komponente	Version	Beschreibung
hive-metastore-server	2.3.2-amzn-0	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.3.2-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.0.1	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
livy-server	0.4.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	0.12.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.
nvidia-cuda	9.0.176	Nvidia-Treiber und Cuda-Toolkit
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.

Komponente	Version	Beschreibung
phoenix-library	4.11.0-HBase-1.3	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.11.0-HBase-1.3	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.187	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.187	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
spark-client	2.2.1	Spark-Befehlszeilen-Clients.
spark-history-server	2.2.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.2.1	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.2.1	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.



Komponente	Version	Beschreibung
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.7.3	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.10	ZooKeeper Befehlszeilenclient.

### 5.11.4 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

#### emr-5.11.4-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.

Klassifizierungen	Beschreibung
flink-conf	Ändert die flink-conf.yaml-Einstellungen.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4 j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.

Klassifizierungen	Beschreibung
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log

Klassifizierungen	Beschreibung
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der Anwendungsumgebung MapReduce .

Klassifizierungen	Beschreibung
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.

Klassifizierungen	Beschreibung
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.

Klassifizierungen	Beschreibung
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.

Klassifizierungen	Beschreibung
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.11.3

### 5.11.3 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

### Informationen zur Anwendungsversion

	emr-5.11.3	emr-5.11.2	emr-5.11.1	emr-5.11.0
AWS SDK for Java	1.11.238	1,11.238	1,11.238	1,11.238
Python	2,7, 3,4	2,7, 3,4	2,7, 3,4	2,7, 3,4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-



	emr-5.11.3	emr-5.11.2	emr-5.11.1	emr-5.11.0
Delta	-	-	-	-
Flink	1.3.2	1.3.2	1.3.2	1.3.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.1	1.3.1	1.3.1
HCatalog	2.3.2	2.3.2	2.3.2	2.3.2
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.3.2	2.3.2	2.3.2	2.3.2
Hudi	-	-	-	-
Hue	4.0.1	4.0.1	4.0.1	4.0.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	0.4.0	0,4,0	0,4,0	0,4,0
MXNet	0.12.0	0.12.0	0.12.0	0.12.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.11.0	4.11.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.187	0.187	0.187	0.187
Spark	2.2.1	2.2.1	2.2.1	2.2.1

	emr-5.11.3	emr-5.11.2	emr-5.11.1	emr-5.11.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

### 5.11.3 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.11.3. Änderungen beziehen sich auf Version 5.11.2.

Erste Version: 18. Juli 2019

Änderungen, Verbesserungen und behobene Probleme

- Das standardmäßige Amazon Linux AMI für Amazon EMR wurde aktualisiert, sodass es wichtige Sicherheitsupdates für den Linux-Kernel enthält, darunter das „TCP SACK Denial of Service Issue“ ([AWS-2019-005](#)).

### 5.11.3 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-*

`amzn-EmrVersion`. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1,0	Amazon SageMaker Spark-SDK
<code>emr-ddb</code>	4.5.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.4.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-s3-dist-cp</code>	2.8.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
<code>emrfs</code>	2.20.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>flink-client</code>	1.3.2	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
<code>ganglia-monitor</code>	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.

Komponente	Version	Beschreibung
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-6	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-6	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-6	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.7.3-amzn-6	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httpfs-server	2.7.3-amzn-6	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.7.3-amzn-6	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider

Komponente	Version	Beschreibung
hadoop-mapred	2.7.3-amzn-6	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-6	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.3-amzn-6	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-6	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.3.1	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.3.1	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.3.1	HBase-Befehlszeilen-Client.
hbase-rest-server	1.3.1	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.3.1	Service, der einen Thrift-Endpunkt für HBase bereitstellt.

Komponente	Version	Beschreibung
hcatalog-client	2.3.2-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.2-amzn-0	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.3.2-amzn-0	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.3.2-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.3.2-amzn-0	Hive-hbase client.
hive-metastore-server	2.3.2-amzn-0	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.3.2-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.0.1	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
livy-server	0.4.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
mahout-client	0.13.0	Bibliothek für Machine Learning.

Komponente	Version	Beschreibung
mxnet	0.12.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.
nvidia-cuda	9.0.176	Nvidia-Treiber und Cuda-Tool kit
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.11.0-HBase-1.3	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.11.0-HBase-1.3	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.187	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.187	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
spark-client	2.2.1	Spark-Befehlszeilen-Clients.

Komponente	Version	Beschreibung
spark-history-server	2.2.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.2.1	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.2.1	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.7.3	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.10	ZooKeeper Befehlszeilenclient.



## 5.11.3 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

### emr-5.11.3-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Ändert die <code>log4j-cli.properties</code> -Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.

Klassifizierungen	Beschreibung
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.

Klassifizierungen	Beschreibung
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.

Klassifizierungen	Beschreibung
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der Anwendungsumgebung MapReduce .
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.

Klassifizierungen	Beschreibung
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.

Klassifizierungen	Beschreibung
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.11.2

### 5.11.2 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)

- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

## Informationen zur Anwendungsversion

	emr-5.11.2	emr-5.11.1	emr-5.11.0	emr-5.10.1
AWS SDK for Java	1.11.238	1,11.238	1,11.238	1.11.221
Python	2,7, 3,4	2,7, 3,4	2,7, 3,4	2,7, 3,4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.3.2	1.3.2	1.3.2	1.3.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.1	1.3.1	1.3.1
HCatalog	2.3.2	2.3.2	2.3.2	2.3.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.3.2	2.3.2	2.3.2	2.3.1
Hudi	-	-	-	-
Hue	4.0.1	4.0.1	4.0.1	4.0.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-



	emr-5.11.2	emr-5.11.1	emr-5.11.0	emr-5.10.1
Livy	0.4.0	0,4,0	0,4,0	0,4,0
MXNet	0.12.0	0.12.0	0.12.0	0.12.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.11.0	4.11.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.187	0.187	0.187	0.187
Spark	2.2.1	2.2.1	2.2.1	2.2.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

## 5.11.2 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.11.2. Änderungen beziehen sich auf Version 5.11.1.

Erste Version: 29. August 2018

Änderungen, Verbesserungen und behobene Probleme

- Diese Version behebt eine potenzielle Schwachstellen.

## 5.11.2 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>aws-sagemaker-spark-sdk</code>	1,0	Amazon SageMaker Spark-SDK
<code>emr-ddb</code>	4.5.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.4.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-s3-dist-cp</code>	2.8.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
<code>emrfs</code>	2.20.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
flink-client	1.3.2	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-6	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-6	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-6	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.7.3-amzn-6	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.

Komponente	Version	Beschreibung
hadoop-https-server	2.7.3-amzn-6	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.7.3-amzn-6	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.7.3-amzn-6	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-6	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resource-manager	2.7.3-amzn-6	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-6	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.3.1	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.3.1	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.

Komponente	Version	Beschreibung
hbase-client	1.3.1	HBase-Befehlszeilen-Client.
hbase-rest-server	1.3.1	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.3.1	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.2-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.2-amzn-0	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.3.2-amzn-0	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.3.2-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.3.2-amzn-0	Hive-hbase client.
hive-metastore-server	2.3.2-amzn-0	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.3.2-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.

Komponente	Version	Beschreibung
hue-server	4.0.1	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
livy-server	0.4.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	0.12.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.
nvidia-cuda	9.0.176	Nvidia-Treiber und Cuda-Toolkit
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.11.0-HBase-1.3	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.11.0-HBase-1.3	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.

Komponente	Version	Beschreibung
presto-coordinator	0.187	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.187	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
spark-client	2.2.1	Spark-Befehlszeilen-Clients.
spark-history-server	2.2.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.2.1	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.2.1	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.7.3	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.

Komponente	Version	Beschreibung
zookeeper-server	3.4.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.10	ZooKeeper Befehlszeilenclient.

## 5.11.2 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

### emr-5.11.2-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink <code>log4j-yarn-session.properties</code> .



Klassifizierungen	Beschreibung
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.

Klassifizierungen	Beschreibung
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.

Klassifizierungen	Beschreibung
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der Anwendungsumgebung MapReduce .
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.

Klassifizierungen	Beschreibung
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.

Klassifizierungen	Beschreibung
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.11.1

### 5.11.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

#### Informationen zur Anwendungsversion

	emr-5.11.1	emr-5.11.0	emr-5.10.1	emr-5.10.0
AWS SDK for Java	1.11.238	1,11.238	1.11.221	1.11.221
Python	2,7, 3,4	2,7, 3,4	2,7, 3,4	2,7, 3,4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.3.2	1.3.2	1.3.2	1.3.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2

	emr-5.11.1	emr-5.11.0	emr-5.10.1	emr-5.10.0
HBase	1.3.1	1.3.1	1.3.1	1.3.1
HCatalog	2.3.2	2.3.2	2.3.1	2.3.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.3.2	2.3.2	2.3.1	2.3.1
Hudi	-	-	-	-
Hue	4.0.1	4.0.1	4.0.1	4.0.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	0.4.0	0,4,0	0,4,0	0,4,0
MXNet	0.12.0	0.12.0	0.12.0	0.12.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.11.0	4.11.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.187	0.187	0.187	0.187
Spark	2.2.1	2.2.1	2.2.0	2.2.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4



	emr-5.11.1	emr-5.11.0	emr-5.10.1	emr-5.10.0
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

### 5.11.1 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon EMR Version 5.11.1. Änderungen beziehen sich auf Amazon-EMR-Version 5.8.0.

Erste Version: 22. Januar 2018

Änderungen, Verbesserungen und behobene Probleme

- Der Amazon-Linux-Kernel des standardmäßigen Amazon-Linux-AMI für Amazon EMR wurde aktualisiert, um Schwachstellen im Zusammenhang mit spekulativer Ausführung zu beheben (CVE-2017-5715, CVE-2017-5753 und CVE-2017-5754). Weitere Informationen finden Sie unter <https://aws.amazon.com/security/security-bulletins/AWS-2018-013/>.

### 5.11.1 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1,0	Amazon SageMaker Spark-SDK
emr-ddb	4.5.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.4.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.8.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.20.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.3.2	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.

Komponente	Version	Beschreibung
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-6	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-6	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-6	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.7.3-amzn-6	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httpfs-server	2.7.3-amzn-6	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.7.3-amzn-6	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.7.3-amzn-6	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-6	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.

Komponente	Version	Beschreibung
hadoop-yarn-resourcemanager	2.7.3-amzn-6	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-6	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.3.1	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.3.1	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.3.1	HBase-Befehlszeilen-Client.
hbase-rest-server	1.3.1	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.3.1	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.2-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.2-amzn-0	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).

Komponente	Version	Beschreibung
hcatalog-webhcat-server	2.3.2-amzn-0	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.3.2-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.3.2-amzn-0	Hive-hbase client.
hive-metastore-server	2.3.2-amzn-0	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.3.2-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.0.1	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
livy-server	0.4.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	0.12.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.
nvidia-cuda	9.0.176	Nvidia-Treiber und Cuda-Toolkit

Komponente	Version	Beschreibung
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.11.0-HBase-1.3	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.11.0-HBase-1.3	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.187	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.187	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
spark-client	2.2.1	Spark-Befehlszeilen-Clients.
spark-history-server	2.2.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.2.1	In-Memory-Ausführungs-Engine für YARN.

Komponente	Version	Beschreibung
spark-yarn-slave	2.2.1	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.7.3	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.10	ZooKeeper Befehlszeilenclient.

### 5.11.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

## emr-5.11.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der capacity-scheduler.xml-Datei in Hadoop.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4 j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.



Klassifizierungen	Beschreibung
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.

Klassifizierungen	Beschreibung
hive-parquet-logging	Ändert die Werte in der <code>parquet-logging.properties</code> -Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von <code>hive-exec-log</code> Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von Hive. <code>llap-daemon-log</code>
hive-log4j2	Ändert die Werte in der <code>hive-log4j2.properties</code> -Datei in Hive.
hive-site	Ändert die Werte in der <code>hive-site.xml</code> -Datei in Hive.
hiveserver2-site	Ändert die Werte in der <code>hiveserver2-site.xml</code> -Datei von <code>Server2</code> in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der <code>httpfs-site.xml</code> -Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der <code>kms-acls.xml</code> -Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der <code>kms-log4j.properties</code> -Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der <code>kms-site.xml</code> -Datei in Hadoop.

Klassifizierungen	Beschreibung
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der Anwendungsumgebung MapReduce .
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.

Klassifizierungen	Beschreibung
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.

Klassifizierungen	Beschreibung
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.11.0

### 5.11.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

## Informationen zur Anwendungsversion

	emr-5.11.0	emr-5.10.1	emr-5.10.0	emr-5.9.1
AWS SDK for Java	1.11.238	1.11.221	1.11.221	1.11.183
Python	2,7, 3,4	2,7, 3,4	2,7, 3,4	Nicht nachverfolgt
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.3.2	1.3.2	1.3.2	1.3.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.1	1.3.1	1.3.1
HCatalog	2.3.2	2.3.1	2.3.1	2.3.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.3.2	2.3.1	2.3.1	2.3.0
Hudi	-	-	-	-
Hue	4.0.1	4.0.1	4.0.1	4.0.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	0.4.0	0,4,0	0,4,0	0,4,0

	emr-5.11.0	emr-5.10.1	emr-5.10.0	emr-5.9.1
MXNet	0.12.0	0.12.0	0.12.0	-
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.11.0	4.11.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.187	0.187	0.187	0.184
Spark	2.2.1	2.2.0	2.2.0	2.2.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.2
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

## 5.11.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon EMR Version 5.11.0. Änderungen beziehen sich auf Version 5.10.0.

### Upgrades

- Hive 2.3.2
- Spark 2.2.1
- SDK für Java 1.11.238



## Neue Features

- Spark
  - Die Einstellung `spark.decommissioning.timeout.threshold` wurde hinzugefügt, was das Verhalten von Spark bei der Außerbetriebnahme bei Verwendung von Spot Instances verbessert. Weitere Informationen finden Sie unter [Verhalten der Außerbetriebnahme von Knoten konfigurieren](#).
  - Die `aws-sagemaker-spark-sdk` Komponente wurde zu Spark hinzugefügt, die Amazon SageMaker Spark und zugehörige Abhängigkeiten für die Spark-Integration mit [Amazon](#) installiert SageMaker. Sie können Amazon SageMaker Spark verwenden, um Spark-Pipelines für maschinelles Lernen (ML) mithilfe von SageMaker Amazon-Stufen zu erstellen. Weitere Informationen finden Sie in der [SageMaker Spark-Readme-Datei](#) zu Apache Spark GitHub und zur [Verwendung von Apache Spark mit Amazon SageMaker](#) im Amazon SageMaker Developer Guide.

## Bekannte Probleme

- MXNet enthält keine OpenCV-Bibliotheken.
- Hive 2.3.2 setzt `hive.compute.query.using.stats=true` standardmäßig. Dies bewirkt, dass Abfragen Daten aus vorhandenen Statistiken und nicht direkt aus den Daten abrufen, da diese unübersichtlich sein können. Wenn Sie beispielsweise über eine Tabelle mit `hive.compute.query.using.stats=true` verfügen und neue Dateien in die Tabelle LOCATION hochladen, gibt die Abfrage `SELECT COUNT(*)` in der Tabelle die Anzahl aus den Statistiken zurück, anstatt die hinzugefügten Zeilen abzurufen.

Um dieses Problem zu umgehen, verwenden Sie den Befehl `ANALYZE TABLE`, um neue Statistiken zu sammeln, oder legen Sie `hive.compute.query.using.stats=false` fest. Weitere Informationen finden Sie unter [Statistics in Hive](#) in der Apache Hive-Dokumentation.

### 5.11.0 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version

sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen myapp-component der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als 2.2-amzn-2 aufgeführt.

Komponente	Version	Beschreibung
aws-sagemaker-spark-sdk	1,0	Amazon SageMaker Spark-SDK
emr-ddb	4.5.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.4.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.8.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.20.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.3.2	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.

Komponente	Version	Beschreibung
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-6	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-6	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-6	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.7.3-amzn-6	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httpfs-server	2.7.3-amzn-6	HTTP-Endpunkt für HDFS-Operationen.

Komponente	Version	Beschreibung
hadoop-kms-server	2.7.3-amzn-6	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.7.3-amzn-6	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-6	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.3-amzn-6	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-6	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.3.1	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.3.1	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.3.1	HBase-Befehlszeilen-Client.

Komponente	Version	Beschreibung
hbase-rest-server	1.3.1	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.3.1	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.2-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.2-amzn-0	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.3.2-amzn-0	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.3.2-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.3.2-amzn-0	Hive-hbase client.
hive-metastore-server	2.3.2-amzn-0	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.3.2-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.

Komponente	Version	Beschreibung
hue-server	4.0.1	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
livy-server	0.4.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	0.12.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.
nvidia-cuda	9.0.176	Nvidia-Treiber und Cuda-Toolkit
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.11.0-HBase-1.3	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.11.0-HBase-1.3	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.

Komponente	Version	Beschreibung
presto-coordinator	0.187	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.187	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
spark-client	2.2.1	Spark-Befehlszeilen-Clients.
spark-history-server	2.2.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.2.1	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.2.1	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.7.3	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.

Komponente	Version	Beschreibung
zookeeper-server	3.4.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.10	ZooKeeper Befehlszeilenclient.

### 5.11.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

#### emr-5.11.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink <code>log4j-yarn-session.properties</code> .



Klassifizierungen	Beschreibung
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.

Klassifizierungen	Beschreibung
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.

Klassifizierungen	Beschreibung
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der Anwendungsumgebung MapReduce .
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.

Klassifizierungen	Beschreibung
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.

Klassifizierungen	Beschreibung
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.10.1

### 5.10.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

#### Informationen zur Anwendungsversion

	emr-5.10.1	emr-5.10.0	emr-5.9.1	emr-5.9.0
AWS SDK for Java	1.11.221	1.11.221	1.11.183	1.11,183
Python	2,7, 3,4	2,7, 3,4	Nicht nachverfolgt	Nicht nachverfolgt
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.3.2	1.3.2	1.3.2	1.3.2

	emr-5.10.1	emr-5.10.0	emr-5.9.1	emr-5.9.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.1	1.3.1	1.3.1
HCatalog	2.3.1	2.3.1	2.3.0	2.3.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.3.1	2.3.1	2.3.0	2.3.0
Hudi	-	-	-	-
Hue	4.0.1	4.0.1	4.0.1	4.0.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	0.4.0	0,4,0	0,4,0	0,4,0
MXNet	0.12.0	0.12.0	-	-
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.11.0	4.11.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.187	0.187	0.184	0.184
Spark	2.2.0	2.2.0	2.2.0	2.2.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-



	emr-5.10.1	emr-5.10.0	emr-5.9.1	emr-5.9.0
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.2	0.7.2
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

### 5.10.1 Versionshinweise

Dies ist eine Patch-Version, um die AWS Signature Version 4-Authentifizierung für Anfragen an Amazon S3 hinzuzufügen. Alle Anwendungen und Komponenten sind dieselben wie in der vorherigen Amazon-EMR-Version.

#### Important

In dieser Release-Version verwendet Amazon EMR AWS Signature Version 4 ausschließlich zur Authentifizierung von Anfragen an Amazon S3. Weitere Informationen finden Sie unter [Was ist neu](#).

### 5.10.1 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-*amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die

Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als 2.2-amzn-2 aufgeführt.

Komponente	Version	Beschreibung
emr-ddb	4.5.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.4.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.7.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.20.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.3.2	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.

Komponente	Version	Beschreibung
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-5	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-5	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-5	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.7.3-amzn-5	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httpfs-server	2.7.3-amzn-5	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.7.3-amzn-5	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.7.3-amzn-5	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-5	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.

Komponente	Version	Beschreibung
hadoop-yarn-resourcemanager	2.7.3-amzn-5	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-5	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.3.1	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.3.1	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.3.1	HBase-Befehlszeilen-Client.
hbase-rest-server	1.3.1	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.3.1	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.1-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.1-amzn-0	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).

Komponente	Version	Beschreibung
hcatalog-webhcat-server	2.3.1-amzn-0	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.3.1-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.3.1-amzn-0	Hive-hbase client.
hive-metastore-server	2.3.1-amzn-0	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.3.1-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.0.1	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
livy-server	0.4.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	0.12.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.
nvidia-cuda	9.0.176	Nvidia-Treiber und Cuda-Toolkit

Komponente	Version	Beschreibung
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.11.0-HBase-1.3	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.11.0-HBase-1.3	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.187	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.187	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
spark-client	2.2.0	Spark-Befehlszeilen-Clients.
spark-history-server	2.2.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.2.0	In-Memory-Ausführungs-Engine für YARN.

Komponente	Version	Beschreibung
spark-yarn-slave	2.2.0	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.7.3	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.10	ZooKeeper Befehlszeilenclient.

### 5.10.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

## emr-5.10.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der capacity-scheduler.xml-Datei in Hadoop.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4 j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.



Klassifizierungen	Beschreibung
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.

Klassifizierungen	Beschreibung
hive-parquet-logging	Ändert die Werte in der <code>parquet-logging.properties</code> -Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von <code>hive-exec-log</code> Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von Hive. <code>llap-daemon-log</code>
hive-log4j2	Ändert die Werte in der <code>hive-log4j2.properties</code> -Datei in Hive.
hive-site	Ändert die Werte in der <code>hive-site.xml</code> -Datei in Hive.
hiveserver2-site	Ändert die Werte in der <code>hiveserver2-site.xml</code> -Datei von <code>Server2</code> in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der <code>httpfs-site.xml</code> -Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der <code>kms-acls.xml</code> -Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der <code>kms-log4j.properties</code> -Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der <code>kms-site.xml</code> -Datei in Hadoop.

Klassifizierungen	Beschreibung
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der Anwendungsumgebung MapReduce .
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.

Klassifizierungen	Beschreibung
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.

Klassifizierungen	Beschreibung
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.10.0

### 5.10.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

## Informationen zur Anwendungsversion

	emr-5.10.0	emr-5.9.1	emr-5.9.0	emr-5.8.3
AWS SDK for Java	1.11.221	1.11.183	1.11,183	1.11.160
Python	2,7, 3,4	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.3.2	1.3.2	1.3.2	1.3.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.1	1.3.1	1.3.1
HCatalog	2.3.1	2.3.0	2.3.0	2.3.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.3.1	2.3.0	2.3.0	2.3.0
Hudi	-	-	-	-
Hue	4.0.1	4.0.1	4.0.1	3.12.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	0.4.0	0,4,0	0,4,0	-

	emr-5.10.0	emr-5.9.1	emr-5.9.0	emr-5.8.3
MXNet	0.12.0	-	-	-
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.11.0	4.11.0
Pig	0.17.0	0.17.0	0.17.0	0.16.0
Presto	0.187	0.184	0.184	0.170
Spark	2.2.0	2.2.0	2.2.0	2.2.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.2	0.7.2	0.7.2
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

## 5.10.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.10.0. Änderungen beziehen sich auf Amazon-EMR-Version 5.9.0.

### Upgrades

- AWS SDK for Java 1,11.221
- Hive 2.3.1
- Presto 0.187



## Neue Features

- Die Kerberos-Authentifizierung wird jetzt unterstützt. Weitere Informationen finden Sie unter [Verwendung der Kerberos-Authentifizierung](#) im Verwaltungshandbuch für Amazon EMR.
- Unterstützung für IAM-Rollen für EMRFS-Anfragen hinzugefügt. Weitere Informationen finden Sie unter [Konfigurieren von IAM-Rollen für EMRFS Requests zu Amazon S3](#) im Verwaltungshandbuch für Amazon EMR.
- Unterstützung für GPU-basierte P2- und P3-Instance-Typen hinzugefügt. Weitere Informationen finden Sie unter [Amazon-EC2-P2-Instances](#) und [Amazon-EC2-P3-Instances](#). NVIDIA-Treiber 384.81 und CUDA-Treiber 9.0.176 sind auf diesen Instance-Typen standardmäßig installiert.
- Unterstützung für [Apache MXNet](#) hinzugefügt.

## Änderungen, Verbesserungen und behobene Probleme

- Presto
  - Unterstützung für die Verwendung des AWS Glue-Datenkatalogs als Standard-Hive-Metastore hinzugefügt. Weitere Informationen finden Sie unter [Presto mit dem AWS Glue-Datenkatalog verwenden](#).
  - Unterstützung von [koordinatenbasierten Funktionen](#) hinzugefügt.
  - Unterstützung eines [Überlaufs auf die Festplatte](#) für Joins hinzugefügt.
  - Unterstützung für den [Redshift-Connector](#) hinzugefügt.
- Spark
  - Rückportierung von [SPARK-20640](#), mit der der rpc-Timeout und die erneuten Versuche für das Shuffling von Registrierungswerten unter Verwendung von `spark.shuffle.registration.timeout` bzw. `spark.shuffle.registration.maxAttempts` konfiguriert werden können.
  - [SPARK-21549](#) wurde rückportiert, wodurch ein Fehler behoben wird, der beim Schreiben von benutzerdefinierten Speicherorten in Nicht-HDFS-Speicherorte auftritt. `OutputFormat`
- Rückportierung von [Hadoop-13270](#)
- Die Bibliotheken Numpy, Scipy und Matplotlib wurden aus der grundlegenden Amazon-EMR-AMI entfernt. Wenn diese Bibliotheken für Ihre Anwendung erforderlich sind, werden sie im folgenden Repository bereitgestellt, sodass Sie sie unter Verwendung einer Bootstrap-Aktion mit `yum install` auf allen Knoten installieren können.

- Die Amazon-EMR-Basis-AMI enthält keine Anwendungs-RPM-Paketen mehr, sodass die RPM-Pakete nicht mehr auf Cluster-Knoten vorhanden sind. Benutzerdefinierte AMIs und die Amazon EMR Basis-AMI von Amazon S3 verweisen jetzt auf das RPM-Paket-Repository in .
- Aufgrund der Einführung von der sekundenweisen Abrechnung in Amazon EC2 ist die Standardeinstellung für Verhalten beim Herunterskalieren jetzt Nach Abschluss der Aufgabe beenden anstelle von Zu einem Zeitpunkt auf der Instance beenden. Weitere Informationen finden Sie unter [Konfigurieren eines Cluster-Scale-Downs](#).

## Bekannte Probleme

- MXNet enthält keine OpenCV-Bibliotheken.
- Hive 2.3.1 setzt `hive.compute.query.using.stats=true` standardmäßig. Dies bewirkt, dass Abfragen Daten aus vorhandenen Statistiken und nicht direkt aus den Daten abrufen, da diese unübersichtlich sein können. Wenn Sie beispielsweise über eine Tabelle mit `hive.compute.query.using.stats=true` verfügen und neue Dateien in die Tabelle LOCATION hochladen, gibt die Abfrage `SELECT COUNT(*)` in der Tabelle die Anzahl aus den Statistiken zurück, anstatt die hinzugefügten Zeilen abzurufen.

Um dieses Problem zu umgehen, verwenden Sie den Befehl `ANALYZE TABLE`, um neue Statistiken zu sammeln, oder legen Sie `hive.compute.query.using.stats=false` fest. Weitere Informationen finden Sie unter [Statistics in Hive](#) in der Apache Hive-Dokumentation.

## 5.10.0 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-`amzn-EmrVersion`. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
emr-ddb	4.5.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.4.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.7.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.20.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.3.2	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den

Komponente	Version	Beschreibung
		Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-5	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-5	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-5	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.7.3-amzn-5	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httpfs-server	2.7.3-amzn-5	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.7.3-amzn-5	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.7.3-amzn-5	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-5	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.

Komponente	Version	Beschreibung
hadoop-yarn-resourcemanager	2.7.3-amzn-5	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-5	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.3.1	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.3.1	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.3.1	HBase-Befehlszeilen-Client.
hbase-rest-server	1.3.1	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.3.1	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.1-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.1-amzn-0	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).

Komponente	Version	Beschreibung
hcatalog-webhcat-server	2.3.1-amzn-0	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.3.1-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.3.1-amzn-0	Hive-hbase client.
hive-metastore-server	2.3.1-amzn-0	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.3.1-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.0.1	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
livy-server	0.4.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
mahout-client	0.13.0	Bibliothek für Machine Learning.
mxnet	0.12.0	Eine flexible, skalierbare und effiziente Bibliothek für Deep Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.
nvidia-cuda	9.0.176	Nvidia-Treiber und Cuda-Toolkit

Komponente	Version	Beschreibung
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.11.0-HBase-1.3	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.11.0-HBase-1.3	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.187	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.187	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
spark-client	2.2.0	Spark-Befehlszeilen-Clients.
spark-history-server	2.2.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.2.0	In-Memory-Ausführungs-Engine für YARN.

Komponente	Version	Beschreibung
spark-yarn-slave	2.2.0	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.7.3	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.10	ZooKeeper Befehlszeilenclient.

### 5.10.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).



## emr-5.10.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der capacity-scheduler.xml-Datei in Hadoop.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4 j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.

Klassifizierungen	Beschreibung
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.

Klassifizierungen	Beschreibung
hive-parquet-logging	Ändert die Werte in der <code>parquet-logging.properties</code> -Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von <code>hive-exec-log</code> Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von Hive. <code>llap-daemon-log</code>
hive-log4j2	Ändert die Werte in der <code>hive-log4j2.properties</code> -Datei in Hive.
hive-site	Ändert die Werte in der <code>hive-site.xml</code> -Datei in Hive.
hiveserver2-site	Ändert die Werte in der <code>hiveserver2-site.xml</code> -Datei von <code>Server2</code> in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der <code>httpfs-site.xml</code> -Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der <code>kms-acls.xml</code> -Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der <code>kms-log4j.properties</code> -Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der <code>kms-site.xml</code> -Datei in Hadoop.

Klassifizierungen	Beschreibung
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der Anwendungsumgebung MapReduce .
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.

Klassifizierungen	Beschreibung
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.

Klassifizierungen	Beschreibung
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.9.1

### 5.9.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

## Informationen zur Anwendungsversion

	emr-5.9.1	emr-5.9.0	emr-5.8.3	emr-5.8.2
AWS SDK for Java	1.11.183	1.11,183	1.11.160	1,11.160
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.3.2	1.3.2	1.3.1	1.3.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.1	1.3.1	1.3.1
HCatalog	2.3.0	2.3.0	2.3.0	2.3.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.3.0	2.3.0	2.3.0	2.3.0
Hudi	-	-	-	-
Hue	4.0.1	4.0.1	3.12.0	3.12.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	0.4.0	0,4,0	-	-



	emr-5.9.1	emr-5.9.0	emr-5.8.3	emr-5.8.2
MXNet	-	-	-	-
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.11.0	4.11.0
Pig	0.17.0	0.17.0	0.16.0	0,16,0
Presto	0.184	0.184	0.170	0.170
Spark	2.2.0	2.2.0	2.2.0	2.2.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.2	0.7.2	0.7.2	0.7.2
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

### 5.9.1 Versionshinweise

Dies ist eine Patch-Version, um die AWS Signature Version 4-Authentifizierung für Anfragen an Amazon S3 hinzuzufügen. Alle Anwendungen und Komponenten sind dieselben wie in der vorherigen Amazon-EMR-Version.

#### Important

In dieser Release-Version verwendet Amazon EMR AWS Signature Version 4 ausschließlich zur Authentifizierung von Anfragen an Amazon S3. Weitere Informationen finden Sie unter [Was ist neu](#).

## 5.9.1 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>emr-ddb</code>	4.4.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.4.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-s3-dist-cp</code>	2.7.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
<code>emrfs</code>	2.19.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
flink-client	1.3.2	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-4	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-4	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-4	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.7.3-amzn-4	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.

Komponente	Version	Beschreibung
hadoop-httfs-server	2.7.3-amzn-4	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.7.3-amzn-4	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.7.3-amzn-4	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-4	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.3-amzn-4	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-4	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.3.1	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.3.1	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.

Komponente	Version	Beschreibung
hbase-client	1.3.1	HBase-Befehlszeilen-Client.
hbase-rest-server	1.3.1	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.3.1	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.0-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.0-amzn-0	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.3.0-amzn-0	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.3.0-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.3.0-amzn-0	Hive-hbase client.
hive-metastore-server	2.3.0-amzn-0	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.3.0-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.

Komponente	Version	Beschreibung
hue-server	4.0.1	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
livy-server	0.4.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
mahout-client	0.13.0	Bibliothek für Machine Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.11.0-HBase-1.3	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.11.0-HBase-1.3	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.184	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.184	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.

Komponente	Version	Beschreibung
spark-client	2.2.0	Spark-Befehlszeilen-Clients.
spark-history-server	2.2.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.2.0	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.2.0	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.7.2	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.10	ZooKeeper Befehlszeilenclient.

## 5.9.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

### emr-5.9.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Ändert die <code>log4j-cli.properties</code> -Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.



Klassifizierungen	Beschreibung
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der proto-hive-site.xml-Datei von HCatalog.
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.

Klassifizierungen	Beschreibung
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.

Klassifizierungen	Beschreibung
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der Anwendungsumgebung MapReduce .
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.

Klassifizierungen	Beschreibung
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.

Klassifizierungen	Beschreibung
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.9.0

### 5.9.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)

- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

## Informationen zur Anwendungsversion

	emr-5.9.0	emr-5.8.3	emr-5.8.2	emr-5.8.1
AWS SDK for Java	1.11.183	1.11.160	1,11.160	1,11.160
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.3.2	1.3.1	1.3.1	1.3.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.1	1.3.1	1.3.1
HCatalog	2.3.0	2.3.0	2.3.0	2.3.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.3.0	2.3.0	2.3.0	2.3.0
Hudi	-	-	-	-
Hue	4.0.1	3.12.0	3.12.0	3.12.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-

	emr-5.9.0	emr-5.8.3	emr-5.8.2	emr-5.8.1
Livy	0.4.0	-	-	-
MXNet	-	-	-	-
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.11.0	4.11.0
Pig	0.17.0	0.16.0	0,16,0	0,16,0
Presto	0.184	0.170	0.170	0.170
Spark	2.2.0	2.2.0	2.2.0	2.2.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.2	0.7.2	0.7.2	0.7.2
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

## 5.9.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.9.0. Änderungen beziehen sich auf Amazon-EMR-Version 5.8.0.

Veröffentlichungsdatum: 5. Oktober 2017

Letzte Aktualisierung der Funktionen: 12. Oktober 2017



## Upgrades

- AWS SDK for Java Ausführung 1.11.183
- Flink 1.3.2
- Hue 4.0.1
- Pig 0.17.0
- Presto 0.184

## Neue Features

- Livy-Support hinzugefügt (Version 0.4.0-incubating). Weitere Informationen finden Sie unter [Apache Livy](#).
- Support für Hue-Notebooks für Spark hinzugefügt.
- Support für die i3-Serie für Amazon-EC2-Instances hinzugefügt (12. Oktober 2017).

## Änderungen, Verbesserungen und behobene Probleme

- Spark
  - Neue Funktionen hinzugefügt, die sicherstellen, dass Spark die Knotenbeendigung infolge einer manuellen Größenänderung oder einer Anforderung einer Auto Scaling-Richtlinie zuverlässiger ausführt. Weitere Informationen finden Sie unter [Verhalten der Außerbetriebnahme von Knoten konfigurieren](#).
  - Anstelle von 3DES wird SSL-Verschlüsselung bei der Übertragung für den Blocktransferservice verwendet, was die Leistung bei der Verwendung von Amazon-EC2-Instance-Typen mit AES-NI verbessert.
  - [SPARK-21494](#) rückportiert.
- Zeppelin
  - [ZEPPELIN-2377](#) rückportiert.
- HBase
  - Patch [HBASE-18533 hinzugefügt, der zusätzliche Werte für die HBase-Konfiguration](#) unter Verwendung der Konfigurationsklassifizierung ermöglicht. BucketCache hbase-site
- Hue
  - AWS Glue Data Catalog-Unterstützung für den Hive-Abfrageeditor in Hue hinzugefügt.

- Hauptbenutzer in Hue können standardmäßig alle auf Dateien zugreifen, auf die Amazon-EMR-IAM-Rollen zugreifen dürfen. Neu erstellte Benutzer verfügen nicht automatisch über Zugriffsberechtigungen auf den Amazon-S3-Filebrowser und für ihre Gruppen müssen die `filebrowser.s3_access`-Berechtigungen aktiviert sein.
- Es wurde ein Problem behoben, das dazu führte, dass auf die zugrunde liegenden JSON-Daten, die mit dem AWS Glue-Datenkatalog erstellt wurden, nicht zugegriffen werden konnte.

## Bekannte Probleme

- Der Cluster-Start schlägt fehl, wenn alle Anwendungen installiert und die Standardgröße des Amazon-EBS-Stamm-Volumes nicht geändert wurde. Um das Problem zu umgehen, verwenden Sie den `aws emr create-cluster` Befehl von AWS CLI und geben Sie einen größeren `--ebs-root-volume-size` Parameter an.
- Hive 2.3.0 setzt `hive.compute.query.using.stats=true` standardmäßig. Dies bewirkt, dass Abfragen Daten aus vorhandenen Statistiken und nicht direkt aus den Daten abrufen, da diese unübersichtlich sein können. Wenn Sie beispielsweise über eine Tabelle mit `hive.compute.query.using.stats=true` verfügen und neue Dateien in die Tabelle LOCATION hochladen, gibt die Abfrage `SELECT COUNT(*)` in der Tabelle die Anzahl aus den Statistiken zurück, anstatt die hinzugefügten Zeilen abzurufen.

Um dieses Problem zu umgehen, verwenden Sie den Befehl `ANALYZE TABLE`, um neue Statistiken zu sammeln, oder legen Sie `hive.compute.query.using.stats=false` fest. Weitere Informationen finden Sie unter [Statistics in Hive](#) in der Apache Hive-Dokumentation.

## 5.9.0 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-`amzn-EmrVersion`. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die

Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als 2.2-amzn-2 aufgeführt.

Komponente	Version	Beschreibung
emr-ddb	4.4.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.4.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.7.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.19.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.3.2	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.

Komponente	Version	Beschreibung
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-4	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-4	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-4	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.7.3-amzn-4	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httpfs-server	2.7.3-amzn-4	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.7.3-amzn-4	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.7.3-amzn-4	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-4	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.

Komponente	Version	Beschreibung
hadoop-yarn-resourcemanager	2.7.3-amzn-4	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-4	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.3.1	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.3.1	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.3.1	HBase-Befehlszeilen-Client.
hbase-rest-server	1.3.1	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.3.1	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.0-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.0-amzn-0	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).

Komponente	Version	Beschreibung
hcatalog-webhcat-server	2.3.0-amzn-0	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.3.0-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.3.0-amzn-0	Hive-hbase client.
hive-metastore-server	2.3.0-amzn-0	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.3.0-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	4.0.1	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
livy-server	0.4.0-incubating	REST-Schnittstelle für die Interaktion mit Apache Spark
mahout-client	0.13.0	Bibliothek für Machine Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.

Komponente	Version	Beschreibung
phoenix-library	4.11.0-HBase-1.3	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.11.0-HBase-1.3	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.184	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.184	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.17.0	Pig-Befehlszeilen-Client.
spark-client	2.2.0	Spark-Befehlszeilen-Clients.
spark-history-server	2.2.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.2.0	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.2.0	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.

Komponente	Version	Beschreibung
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.7.2	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.10	ZooKeeper Befehlszeilenclient.

## 5.9.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

### emr-5.9.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.



Klassifizierungen	Beschreibung
flink-conf	Ändert die flink-conf.yaml-Einstellungen.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4 j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.

Klassifizierungen	Beschreibung
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log

Klassifizierungen	Beschreibung
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
livy-conf	Ändert die Werte in der livy.conf-Datei von Livy.
livy-env	Ändert die Werte in der Livy-Umgebung.
livy-log4j	Ändert die log4j.properties-Einstellungen für Livy.
mapred-env	Ändern Sie die Werte in der Anwendungsumgebung MapReduce .

Klassifizierungen	Beschreibung
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.

Klassifizierungen	Beschreibung
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.

Klassifizierungen	Beschreibung
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.

Klassifizierungen	Beschreibung
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.8.3

### 5.8.3 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

#### Informationen zur Anwendungsversion

	emr-5.8.3	emr-5.8.2	emr-5.8.1	emr-5.8.0
AWS SDK for Java	1.11.160	1,11.160	1,11.160	1,11.160
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	2.11.8	2.11.8	2.11.8	2.11.8

	emr-5.8.3	emr-5.8.2	emr-5.8.1	emr-5.8.0
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.3.1	1.3.1	1.3.1	1.3.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.1	1.3.1	1.3.1
HCatalog	2.3.0	2.3.0	2.3.0	2.3.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.3.0	2.3.0	2.3.0	2.3.0
Hudi	-	-	-	-
Hue	3.12.0	3.12.0	3.12.0	3.12.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.11.0	4.11.0
Pig	0.16.0	0,16,0	0,16,0	0,16,0



	emr-5.8.3	emr-5.8.2	emr-5.8.1	emr-5.8.0
Presto	0.170	0.170	0.170	0.170
Spark	2.2.0	2.2.0	2.2.0	2.2.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.2	0.7.2	0.7.2	0.7.2
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

### 5.8.3 Versionshinweise

Dies ist eine Patch-Version, um die AWS Signature Version 4-Authentifizierung für Anfragen an Amazon S3 hinzuzufügen. Alle Anwendungen und Komponenten sind dieselben wie in der vorherigen Amazon-EMR-Version.

#### Important

In dieser Release-Version verwendet Amazon EMR AWS Signature Version 4 ausschließlich zur Authentifizierung von Anfragen an Amazon S3. Weitere Informationen finden Sie unter [Was ist neu](#).

### 5.8.3 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version

sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen myapp-component der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als 2.2-amzn-2 aufgeführt.

Komponente	Version	Beschreibung
emr-ddb	4.4.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.4.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.6.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.18.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.3.1	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem

Komponente	Version	Beschreibung
		zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-3	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-3	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-3	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.7.3-amzn-3	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httpfs-server	2.7.3-amzn-3	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.7.3-amzn-3	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider

Komponente	Version	Beschreibung
hadoop-mapred	2.7.3-amzn-3	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-3	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.3-amzn-3	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-3	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.3.1	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.3.1	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.3.1	HBase-Befehlszeilen-Client.
hbase-rest-server	1.3.1	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.3.1	Service, der einen Thrift-Endpunkt für HBase bereitstellt.

Komponente	Version	Beschreibung
hcatalog-client	2.3.0-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.0-amzn-0	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.3.0-amzn-0	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.3.0-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.3.0-amzn-0	Hive-hbase client.
hive-metastore-server	2.3.0-amzn-0	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.3.0-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.12.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.13.0	Bibliothek für Machine Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.

Komponente	Version	Beschreibung
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.11.0-HBase-1.3	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.11.0-HBase-1.3	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.170	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.170	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.16.0-amzn-1	Pig-Befehlszeilen-Client.
spark-client	2.2.0	Spark-Befehlszeilen-Clients.
spark-history-server	2.2.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.2.0	In-Memory-Ausführungs-Engine für YARN.

Komponente	Version	Beschreibung
spark-yarn-slave	2.2.0	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.7.2	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.10	ZooKeeper Befehlszeilenclient.

### 5.8.3 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

## emr-5.8.3-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der capacity-scheduler.xml-Datei in Hadoop.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4 j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.



Klassifizierungen	Beschreibung
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.

Klassifizierungen	Beschreibung
hive-parquet-logging	Ändert die Werte in der <code>parquet-logging.properties</code> -Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von <code>hive-exec-log</code> Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von Hive. <code>llap-daemon-log</code>
hive-log4j2	Ändert die Werte in der <code>hive-log4j2.properties</code> -Datei in Hive.
hive-site	Ändert die Werte in der <code>hive-site.xml</code> -Datei in Hive.
hiveserver2-site	Ändert die Werte in der <code>hiveserver2-site.xml</code> -Datei von <code>Server2</code> in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der <code>httpfs-site.xml</code> -Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der <code>kms-acls.xml</code> -Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der <code>kms-log4j.properties</code> -Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der <code>kms-site.xml</code> -Datei in Hadoop.

Klassifizierungen	Beschreibung
mapred-env	Ändern Sie die Werte in der Anwendungsumgebung MapReduce .
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.

Klassifizierungen	Beschreibung
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.8.2

### 5.8.2 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

### Informationen zur Anwendungsversion

	emr-5.8.2	emr-5.8.1	emr-5.8.0	emr-5.7.1
AWS SDK for Java	1.11.160	1,11.160	1,11.160	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt

	emr-5.8.2	emr-5.8.1	emr-5.8.0	emr-5.7.1
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.3.1	1.3.1	1.3.1	1.3.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.1	1.3.1	1.3.1
HCatalog	2.3.0	2.3.0	2.3.0	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.3.0	2.3.0	2.3.0	2.1.1
Hudi	-	-	-	-
Hue	3.12.0	3.12.0	3.12.0	3.12.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.11.0	4.11.0

	emr-5.8.2	emr-5.8.1	emr-5.8.0	emr-5.7.1
Pig	0.16.0	0,16,0	0,16,0	0,16,0
Presto	0.170	0.170	0.170	0.170
Spark	2.2.0	2.2.0	2.2.0	2.1.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.2	0.7.2	0.7.2	0.7.2
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

## 5.8.2 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.8.2. Änderungen beziehen sich auf Version 5.8.1.

Erste Version: 29. März 2018

Änderungen, Verbesserungen und behobene Probleme

- Der Amazon-Linux-Kernel des standardmäßigen Amazon-Linux-AMI für Amazon EMR wurde aktualisiert, um potenzielle Schwachstellen zu schließen.

## 5.8.2 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version



sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen myapp-component der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als 2.2-amzn-2 aufgeführt.

Komponente	Version	Beschreibung
emr-ddb	4.4.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.4.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.6.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.18.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.3.1	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem

Komponente	Version	Beschreibung
		zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-3	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-3	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-3	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.7.3-amzn-3	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httpfs-server	2.7.3-amzn-3	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.7.3-amzn-3	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider

Komponente	Version	Beschreibung
hadoop-mapred	2.7.3-amzn-3	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-3	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.3-amzn-3	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-3	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.3.1	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.3.1	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.3.1	HBase-Befehlszeilen-Client.
hbase-rest-server	1.3.1	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.3.1	Service, der einen Thrift-Endpunkt für HBase bereitstellt.

Komponente	Version	Beschreibung
hcatalog-client	2.3.0-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.0-amzn-0	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.3.0-amzn-0	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.3.0-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.3.0-amzn-0	Hive-hbase client.
hive-metastore-server	2.3.0-amzn-0	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.3.0-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.12.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.13.0	Bibliothek für Machine Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.

Komponente	Version	Beschreibung
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.11.0-HBase-1.3	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.11.0-HBase-1.3	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.170	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.170	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.16.0-amzn-1	Pig-Befehlszeilen-Client.
spark-client	2.2.0	Spark-Befehlszeilen-Clients.
spark-history-server	2.2.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.2.0	In-Memory-Ausführungs-Engine für YARN.

Komponente	Version	Beschreibung
spark-yarn-slave	2.2.0	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.7.2	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.10	ZooKeeper Befehlszeilenclient.

## 5.8.2 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

## emr-5.8.2-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der capacity-scheduler.xml-Datei in Hadoop.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4 j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.

Klassifizierungen	Beschreibung
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.



Klassifizierungen	Beschreibung
hive-parquet-logging	Ändert die Werte in der <code>parquet-logging.properties</code> -Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von <code>hive-exec-log</code> Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von Hive. <code>llap-daemon-log</code>
hive-log4j2	Ändert die Werte in der <code>hive-log4j2.properties</code> -Datei in Hive.
hive-site	Ändert die Werte in der <code>hive-site.xml</code> -Datei in Hive.
hiveserver2-site	Ändert die Werte in der <code>hiveserver2-site.xml</code> -Datei von <code>Server2</code> in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der <code>httpfs-site.xml</code> -Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der <code>kms-acls.xml</code> -Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der <code>kms-log4j.properties</code> -Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der <code>kms-site.xml</code> -Datei in Hadoop.

Klassifizierungen	Beschreibung
mapred-env	Ändern Sie die Werte in der Anwendungsumgebung MapReduce .
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.

Klassifizierungen	Beschreibung
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.8.1

### 5.8.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

### Informationen zur Anwendungsversion

	emr-5.8.1	emr-5.8.0	emr-5.7.1	emr-5.7.0
AWS SDK for Java	1.11.160	1,11.160	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt

	emr-5.8.1	emr-5.8.0	emr-5.7.1	emr-5.7.0
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.3.1	1.3.1	1.3.0	1.3.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.1	1.3.1	1.3.1
HCatalog	2.3.0	2.3.0	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.3.0	2.3.0	2.1.1	2.1.1
Hudi	-	-	-	-
Hue	3.12.0	3.12.0	3.12.0	3.12.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.11.0	4.11.0

	emr-5.8.1	emr-5.8.0	emr-5.7.1	emr-5.7.0
Pig	0.16.0	0,16,0	0,16,0	0,16,0
Presto	0.170	0.170	0.170	0.170
Spark	2.2.0	2.2.0	2.1.1	2.1.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.2	0.7.2	0.7.2	0.7.2
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

### 5.8.1 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.8.1. Änderungen beziehen sich auf Amazon-EMR-Version 5.8.0.

Erste Version: 22. Januar 2018

Änderungen, Verbesserungen und behobene Probleme

- Der Amazon-Linux-Kernel des standardmäßigen Amazon-Linux-AMI für Amazon EMR wurde aktualisiert, um Schwachstellen im Zusammenhang mit spekulativer Ausführung zu beheben (CVE-2017-5715, CVE-2017-5753 und CVE-2017-5754). Weitere Informationen finden Sie unter <https://aws.amazon.com/security/security-bulletins/AWS-2018-013/>.

### 5.8.1 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der

Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-amzn-*EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>emr-ddb</code>	4.4.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.4.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-s3-dist-cp</code>	2.6.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
<code>emrfs</code>	2.18.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>flink-client</code>	1.3.1	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
<code>ganglia-monitor</code>	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem



Komponente	Version	Beschreibung
		zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-3	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-3	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-3	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.7.3-amzn-3	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httpfs-server	2.7.3-amzn-3	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.7.3-amzn-3	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider

Komponente	Version	Beschreibung
hadoop-mapred	2.7.3-amzn-3	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-3	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.3-amzn-3	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-3	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.3.1	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.3.1	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.3.1	HBase-Befehlszeilen-Client.
hbase-rest-server	1.3.1	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.3.1	Service, der einen Thrift-Endpunkt für HBase bereitstellt.

Komponente	Version	Beschreibung
hcatalog-client	2.3.0-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.0-amzn-0	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.3.0-amzn-0	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.3.0-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.3.0-amzn-0	Hive-hbase client.
hive-metastore-server	2.3.0-amzn-0	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.3.0-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.12.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.13.0	Bibliothek für Machine Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.

Komponente	Version	Beschreibung
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.11.0-HBase-1.3	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.11.0-HBase-1.3	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.170	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.170	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.16.0-amzn-1	Pig-Befehlszeilen-Client.
spark-client	2.2.0	Spark-Befehlszeilen-Clients.
spark-history-server	2.2.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.2.0	In-Memory-Ausführungs-Engine für YARN.

Komponente	Version	Beschreibung
spark-yarn-slave	2.2.0	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.7.2	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.10	ZooKeeper Befehlszeilenclient.

### 5.8.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

## emr-5.8.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der capacity-scheduler.xml-Datei in Hadoop.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4 j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.

Klassifizierungen	Beschreibung
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.

Klassifizierungen	Beschreibung
hive-parquet-logging	Ändert die Werte in der <code>parquet-logging.properties</code> -Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von <code>hive-exec-log</code> Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von Hive. <code>llap-daemon-log</code>
hive-log4j2	Ändert die Werte in der <code>hive-log4j2.properties</code> -Datei in Hive.
hive-site	Ändert die Werte in der <code>hive-site.xml</code> -Datei in Hive.
hiveserver2-site	Ändert die Werte in der <code>hiveserver2-site.xml</code> -Datei von <code>Server2</code> in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der <code>httpfs-site.xml</code> -Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der <code>kms-acls.xml</code> -Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der <code>kms-log4j.properties</code> -Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der <code>kms-site.xml</code> -Datei in Hadoop.



Klassifizierungen	Beschreibung
mapred-env	Ändern Sie die Werte in der Anwendungsumgebung MapReduce .
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.

Klassifizierungen	Beschreibung
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.8.0

### 5.8.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

### Informationen zur Anwendungsversion

	emr-5.8.0	emr-5.7.1	emr-5.7.0	emr-5.6.1
AWS SDK for Java	1.11.160	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt

	emr-5.8.0	emr-5.7.1	emr-5.7.0	emr-5.6.1
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.3.1	1.3.0	1.3.0	1.2.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.1	1.3.1	1.3.0
HCatalog	2.3.0	2.1.1	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.3.0	2.1.1	2.1.1	2.1.1
Hudi	-	-	-	-
Hue	3.12.0	3.12.0	3.12.0	3.12.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.11.0	4.9.0

	emr-5.8.0	emr-5.7.1	emr-5.7.0	emr-5.6.1
Pig	0.16.0	0,16,0	0,16,0	0,16,0
Presto	0.170	0.170	0.170	0.170
Spark	2.2.0	2.1.1	2.1.1	2.1.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.2	0.7.2	0.7.2	0.7.1
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

## 5.8.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.8.0. Änderungen beziehen sich auf Amazon-EMR-Version 5.7.0.

Erste Version: 10. August 2017

Letzte Aktualisierung der Funktionen: 25. September 2017

### Upgrades

- AWS SDK 1.11.160
- Flink 1.3.1
- Hive 2.3.0. Weitere Informationen finden Sie in den [Versionshinweisen](#) auf der Apache Hive-Website.
- Spark 2.2.0. Weitere Informationen finden Sie in den [Versionshinweisen](#) auf der Apache Spark-Website.

## Neue Features

- Support zur Anzeige des Anwendungsverlaufs hinzugefügt (25. September 2017). Weitere Informationen finden Sie unter [Anzeigen des Anwendungsverlaufs](#) im Verwaltungshandbuch für Amazon EMR.

## Änderungen, Verbesserungen und behobene Probleme

- Integration mit dem AWS Glue-Datenkatalog
  - Hive und Spark SQL können jetzt AWS Glue Data Catalog als Hive-Metadatenpeicher verwenden. Weitere Informationen finden Sie unter [Verwenden des AWS Glue-Datenkatalogs als Metastore für Hive](#) und [Verwenden Sie den AWS Glue-Datenkatalog als Metastore für Spark SQL](#).
- Neuen Anwendungsverlauf mit Cluster-Details hinzugefügt, über den Verlaufsdaten für YARN-Anwendungen und zusätzliche Details zu Spark-Anwendungen angezeigt werden können. Weitere Informationen finden Sie unter [Anwendungsverlauf anzeigen](#) im Verwaltungshandbuch für Amazon EMR.
- Oozie
  - [OOZIE-2748](#) wurde rückportiert.
- Hue
  - [HUE-5859](#) wurde rückportiert.
- HBase
  - Patch zum Verfügbarmachen der Startzeit des HBase-Master-Servers über Java Management Extensions (JMX) mithilfe von `getMasterInitializedTime` hinzugefügt.
  - Patch zur Verbesserung der Cluster-Startzeit hinzugefügt.

## Bekannte Probleme

- Der Cluster-Start schlägt fehl, wenn alle Anwendungen installiert und die Standardgröße des Amazon-EBS-Stamm-Volumes nicht geändert wurde. Um das Problem zu umgehen, verwenden Sie den `aws emr create-cluster` Befehl von AWS CLI und geben Sie einen größeren Parameter an. `--ebs-root-volume-size`
- Hive 2.3.0 setzt `hive.compute.query.using.stats=true` standardmäßig. Dies bewirkt, dass Abfragen Daten aus vorhandenen Statistiken und nicht direkt aus den Daten abrufen, da diese unübersichtlich sein können. Wenn Sie beispielsweise über eine Tabelle

mit `hive.compute.query.using.stats=true` verfügen und neue Dateien in die Tabelle `LOCATION` hochladen, gibt die Abfrage `SELECT COUNT(*)` in der Tabelle die Anzahl aus den Statistiken zurück, anstatt die hinzugefügten Zeilen abzurufen.

Um dieses Problem zu umgehen, verwenden Sie den Befehl `ANALYZE TABLE`, um neue Statistiken zu sammeln, oder legen Sie `hive.compute.query.using.stats=false` fest. Weitere Informationen finden Sie unter [Statistics in Hive](#) in der Apache Hive-Dokumentation.

- Spark – Bei der Verwendung von Spark tritt bei dem `appusher`-Daemon bei Spark-Aufträgen mit langer Ausführungsdauer nach einigen Stunden oder Tagen möglicherweise ein Speicherleck-Problem bei dem Datei-Handler auf. Um das Problem zu beheben, stellen Sie eine Verbindung mit dem Primärknoten her und geben `sudo /etc/init.d/appusher stop` ein. Dies beendet den `appusher`-Daemon, den Amazon EMR automatisch neu startet.
- Application history
  - Es sind keine Verlaufsdaten für beendete Spark-Executors verfügbar.
  - Der Anwendungsverlauf ist nicht für Cluster verfügbar, die eine Sicherheitskonfiguration verwenden, um die Verschlüsselung bei der Übertragung zu aktivieren.

## 5.8.0 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.



Komponente	Version	Beschreibung
emr-ddb	4.4.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.4.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.4.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.6.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.18.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.3.1	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den

Komponente	Version	Beschreibung
		Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-3	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-3	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-3	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.7.3-amzn-3	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httfs-server	2.7.3-amzn-3	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.7.3-amzn-3	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.7.3-amzn-3	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-3	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.

Komponente	Version	Beschreibung
hadoop-yarn-resourcemanager	2.7.3-amzn-3	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-3	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.3.1	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.3.1	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.3.1	HBase-Befehlszeilen-Client.
hbase-rest-server	1.3.1	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.3.1	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.3.0-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.3.0-amzn-0	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).

Komponente	Version	Beschreibung
hcatalog-webhcat-server	2.3.0-amzn-0	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.3.0-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.3.0-amzn-0	Hive-hbase client.
hive-metastore-server	2.3.0-amzn-0	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.3.0-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.12.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.13.0	Bibliothek für Machine Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.11.0-HBase-1.3	Die Phoenix-Bibliotheken für den Server und den Client

Komponente	Version	Beschreibung
phoenix-query-server	4.11.0-HBase-1.3	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.170	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.170	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.16.0-amzn-1	Pig-Befehlszeilen-Client.
spark-client	2.2.0	Spark-Befehlszeilen-Clients.
spark-history-server	2.2.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.2.0	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.2.0	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.

Komponente	Version	Beschreibung
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.7.2	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.10	ZooKeeper Befehlszeilenclient.

## 5.8.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

### emr-5.8.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.

Klassifizierungen	Beschreibung
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4 j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.

Klassifizierungen	Beschreibung
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-props-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. props-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.



Klassifizierungen	Beschreibung
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie die Werte in der Anwendungsumgebung MapReduce .
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.

Klassifizierungen	Beschreibung
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.

Klassifizierungen	Beschreibung
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.7.1

### 5.7.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

#### Informationen zur Anwendungsversion

	emr-5.7.1	emr-5.7.0	emr-5.6.1	emr-5.6.0
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.3.0	1.3.0	1.2.1	1.2.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.1	1.3.0	1.3.0
HCatalog	2.1.1	2.1.1	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.1	2.1.1

	emr-5.7.1	emr-5.7.0	emr-5.6.1	emr-5.6.0
Hudi	-	-	-	-
Hue	3.12.0	3.12.0	3.12.0	3.12.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.9.0	4.9.0
Pig	0.16.0	0,16,0	0,16,0	0,16,0
Presto	0.170	0.170	0.170	0.170
Spark	2.1.1	2.1.1	2.1.1	2.1.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.2	0.7.2	0.7.1	0.7.1
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

## 5.7.1 Versionshinweise

Dies ist eine Patch-Version, um die AWS Signature Version 4-Authentifizierung für Anfragen an Amazon S3 hinzuzufügen. Alle Anwendungen und Komponenten sind dieselben wie in der vorherigen Amazon-EMR-Version.

### Important

In dieser Release-Version verwendet Amazon EMR AWS Signature Version 4 ausschließlich zur Authentifizierung von Anfragen an Amazon S3. Weitere Informationen finden Sie unter [Was ist neu](#).

## 5.7.1 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>emr-ddb</code>	4.3.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.3.0	Praktische Bibliotheken für das Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-kinesis	3.3.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.5.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.18.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.3.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-2	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".



Komponente	Version	Beschreibung
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-2	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httpfs-server	2.7.3-amzn-2	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.7.3-amzn-2	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.7.3-amzn-2	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resource-manager	2.7.3-amzn-2	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.

Komponente	Version	Beschreibung
hadoop-yarn-timeline-server	2.7.3-amzn-2	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.3.1	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.3.1	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.3.1	HBase-Befehlszeilen-Client.
hbase-rest-server	1.3.1	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.3.1	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.1.1-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.1.1-amzn-0	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.1.1-amzn-0	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.

Komponente	Version	Beschreibung
hive-client	2.1.1-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.1.1-amzn-0	Hive-hbase client.
hive-metastore-server	2.1.1-amzn-0	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.1.1-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.12.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.13.0	Bibliothek für Machine Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.11.0-HBase-1.3	Die Phoenix-Bibliotheken für den Server und den Client

Komponente	Version	Beschreibung
phoenix-query-server	4.11.0-HBase-1.3	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.170	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.170	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.16.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	2.1.1	Spark-Befehlszeilen-Clients.
spark-history-server	2.1.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.1.1	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.1.1	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.

Komponente	Version	Beschreibung
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.7.2	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.10	ZooKeeper Befehlszeilenclient.

## 5.7.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

### emr-5.7.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.

Klassifizierungen	Beschreibung
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4 j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.

Klassifizierungen	Beschreibung
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.

Klassifizierungen	Beschreibung
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie die Werte in der Anwendungsumgebung MapReduce .
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.



Klassifizierungen	Beschreibung
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.

Klassifizierungen	Beschreibung
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.7.0

### 5.7.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

#### Informationen zur Anwendungsversion

	emr-5.7.0	emr-5.6.1	emr-5.6.0	emr-5.5.4
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.3.0	1.2.1	1.2.1	1.2.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.0	1.3.0	1.3.0
HCatalog	2.1.1	2.1.1	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.1	2.1.1

	emr-5.7.0	emr-5.6.1	emr-5.6.0	emr-5.5.4
Hudi	-	-	-	-
Hue	3.12.0	3.12.0	3.12.0	3.12.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.13.0	0.13.0	0.13.0	0.12.2
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.9.0	4.9.0	4.9.0
Pig	0.16.0	0,16,0	0,16,0	0,16,0
Presto	0.170	0.170	0.170	0.170
Spark	2.1.1	2.1.1	2.1.1	2.1.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.2	0.7.1	0.7.1	0.7.1
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

## 5.7.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon EMR Version 5.7.0. Änderungen beziehen sich auf Amazon-EMR-Version 5.6.0.

Veröffentlichungsdatum: 13. Juli 2017

### Upgrades

- Flink 1.3.0
- Phoenix 4.11.0
- Zeppelin 0.7.2

### Neue Features

- Die Möglichkeit zum Angeben eines benutzerdefinierten Amazon Linux AMI beim Erstellen eines Clusters wurde hinzugefügt. Weitere Informationen finden Sie unter [Verwenden eines benutzerdefinierten AMI](#).

### Änderungen, Verbesserungen und behobene Probleme

- HBase
  - Die Möglichkeit zum Konfigurieren von HBase Read Replica-Cluster wurde hinzugefügt. Weitere Informationen finden Sie unter [Verwenden eines Read Replica-Clusters](#)
  - Mehrere Fehlerbehebungen und Verbesserungen
- Presto – neue Konfigurationsfunktion `node.properties` hinzugefügt.
- YARN – neue Konfigurationsfunktion `container-log4j.properties` hinzugefügt
- Sqoop – [SQOOP-2880](#) rückportiert und damit einen Parameter eingeführt, mit dem Sie das temporäre Sqoop-Verzeichnis festlegen können.

## 5.7.0 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version

sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen myapp-component der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als 2.2-amzn-2 aufgeführt.

Komponente	Version	Beschreibung
emr-ddb	4.3.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.3.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.3.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.5.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.18.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.3.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem

Komponente	Version	Beschreibung
		zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-2	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-2	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httpfs-server	2.7.3-amzn-2	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.7.3-amzn-2	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider



Komponente	Version	Beschreibung
hadoop-mapred	2.7.3-amzn-2	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.3.1	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.3.1	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.3.1	HBase-Befehlszeilen-Client.
hbase-rest-server	1.3.1	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.3.1	Service, der einen Thrift-Endpunkt für HBase bereitstellt.

Komponente	Version	Beschreibung
hcatalog-client	2.1.1-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.1.1-amzn-0	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.1.1-amzn-0	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.1.1-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.1.1-amzn-0	Hive-hbase client.
hive-metastore-server	2.1.1-amzn-0	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.1.1-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.12.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.13.0	Bibliothek für Machine Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.

Komponente	Version	Beschreibung
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.11.0-HBase-1.3	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.11.0-HBase-1.3	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.170	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.170	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.16.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	2.1.1	Spark-Befehlszeilen-Clients.
spark-history-server	2.1.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.1.1	In-Memory-Ausführungs-Engine für YARN.

Komponente	Version	Beschreibung
spark-yarn-slave	2.1.1	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.7.2	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.10	ZooKeeper Befehlszeilenclient.

## 5.7.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

## emr-5.7.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der capacity-scheduler.xml-Datei in Hadoop.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4 j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.

Klassifizierungen	Beschreibung
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.

Klassifizierungen	Beschreibung
hive-parquet-logging	Ändert die Werte in der <code>parquet-logging.properties</code> -Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von <code>hive-exec-log</code> Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von Hive. <code>llap-daemon-log</code>
hive-log4j2	Ändert die Werte in der <code>hive-log4j2.properties</code> -Datei in Hive.
hive-site	Ändert die Werte in der <code>hive-site.xml</code> -Datei in Hive.
hiveserver2-site	Ändert die Werte in der <code>hiveserver2-site.xml</code> -Datei von <code>Server2</code> in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der <code>httpfs-site.xml</code> -Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der <code>kms-acls.xml</code> -Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der <code>kms-log4j.properties</code> -Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der <code>kms-site.xml</code> -Datei in Hadoop.

Klassifizierungen	Beschreibung
mapred-env	Ändern Sie die Werte in der Anwendungsumgebung MapReduce .
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.



Klassifizierungen	Beschreibung
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.

Klassifizierungen	Beschreibung
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.6.1

### 5.6.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

### Informationen zur Anwendungsversion

	emr-5.6.1	emr-5.6.0	emr-5.5.4	emr-5.5.3
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt

	emr-5.6.1	emr-5.6.0	emr-5.5.4	emr-5.5.3
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.2.1	1.2.1	1.2.0	1.2.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.0	1.3.0	1.3.0	1.3.0
HCatalog	2.1.1	2.1.1	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.1	2.1.1
Hudi	-	-	-	-
Hue	3.12.0	3.12.0	3.12.0	3.12.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.13.0	0.13.0	0.12.2	0.12.2
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.9.0	4.9.0	4.9.0	4.9.0

	emr-5.6.1	emr-5.6.0	emr-5.5.4	emr-5.5.3
Pig	0.16.0	0,16,0	0,16,0	0,16,0
Presto	0.170	0.170	0.170	0.170
Spark	2.1.1	2.1.1	2.1.0	2.1.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.1	0.7.1	0.7.1	0.7.1
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

### 5.6.1 Versionshinweise

Dies ist eine Patch-Version, um die AWS Signature Version 4-Authentifizierung für Anfragen an Amazon S3 hinzuzufügen. Alle Anwendungen und Komponenten sind dieselben wie in der vorherigen Amazon-EMR-Version.

#### Important

In dieser Release-Version verwendet Amazon EMR AWS Signature Version 4 ausschließlich zur Authentifizierung von Anfragen an Amazon S3. Weitere Informationen finden Sie unter [Was ist neu](#).

### 5.6.1 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der

Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-amzn-*EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>emr-ddb</code>	4.3.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.3.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.3.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-s3-dist-cp</code>	2.5.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
<code>emrfs</code>	2.17.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>flink-client</code>	1.2.1	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
<code>ganglia-monitor</code>	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem

Komponente	Version	Beschreibung
		zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-2	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-2	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httpfs-server	2.7.3-amzn-2	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.7.3-amzn-2	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider

Komponente	Version	Beschreibung
hadoop-mapred	2.7.3-amzn-2	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.3.1	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.3.1	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.3.1	HBase-Befehlszeilen-Client.
hbase-rest-server	1.3.1	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.3.1	Service, der einen Thrift-Endpunkt für HBase bereitstellt.



Komponente	Version	Beschreibung
hcatalog-client	2.1.1-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.1.1-amzn-0	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.1.1-amzn-0	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.1.1-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.1.1-amzn-0	Hive-hbase client.
hive-metastore-server	2.1.1-amzn-0	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.1.1-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.12.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.13.0	Bibliothek für Machine Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.

Komponente	Version	Beschreibung
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.9.0-HBase-1.2	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.9.0-HBase-1.2	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.170	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.170	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.16.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	2.1.1	Spark-Befehlszeilen-Clients.
spark-history-server	2.1.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.1.1	In-Memory-Ausführungs-Engine für YARN.

Komponente	Version	Beschreibung
spark-yarn-slave	2.1.1	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.7.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.10	ZooKeeper Befehlszeilenclient.

### 5.6.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

## emr-5.6.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der capacity-scheduler.xml-Datei in Hadoop.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4 j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.

Klassifizierungen	Beschreibung
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.

Klassifizierungen	Beschreibung
hive-parquet-logging	Ändert die Werte in der <code>parquet-logging.properties</code> -Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von <code>hive-exec-log</code> Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von Hive. <code>llap-daemon-log</code>
hive-log4j2	Ändert die Werte in der <code>hive-log4j2.properties</code> -Datei in Hive.
hive-site	Ändert die Werte in der <code>hive-site.xml</code> -Datei in Hive.
hiveserver2-site	Ändert die Werte in der <code>hiveserver2-site.xml</code> -Datei von <code>Server2</code> in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der <code>httpfs-site.xml</code> -Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der <code>kms-acls.xml</code> -Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der <code>kms-log4j.properties</code> -Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der <code>kms-site.xml</code> -Datei in Hadoop.

Klassifizierungen	Beschreibung
mapred-env	Ändern Sie die Werte in der Anwendungsumgebung MapReduce .
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.



Klassifizierungen	Beschreibung
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.

Klassifizierungen	Beschreibung
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.6.0

### 5.6.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

### Informationen zur Anwendungsversion

	emr-5.6.0	emr-5.5.4	emr-5.5.3	emr-5.5.2
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt

	emr-5.6.0	emr-5.5.4	emr-5.5.3	emr-5.5.2
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.2.1	1.2.0	1.2.0	1.2.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.0	1.3.0	1.3.0	1.3.0
HCatalog	2.1.1	2.1.1	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.1	2.1.1
Hudi	-	-	-	-
Hue	3.12.0	3.12.0	3.12.0	3.12.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.13.0	0.12.2	0.12.2	0.12.2
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.9.0	4.9.0	4.9.0	4.9.0

	emr-5.6.0	emr-5.5.4	emr-5.5.3	emr-5.5.2
Pig	0.16.0	0,16,0	0,16,0	0,16,0
Presto	0.170	0.170	0.170	0.170
Spark	2.1.1	2.1.0	2.1.0	2.1.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.1	0.7.1	0.7.1	0.7.1
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

## 5.6.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.6.0. Änderungen beziehen sich auf Amazon-EMR-Version 5.5.0.

Veröffentlichungsdatum: 5. Juni 2017

### Upgrades

- Flink 1.2.1
- HBase 1.3.1
- Mahout 0.13.0. Dies ist die erste Version von Mahout, die Spark 2.x in Amazon-EMR-Version 5.0 und höher unterstützt.
- Spark 2.1.1

### Änderungen, Verbesserungen und behobene Probleme

- Presto

- Die Möglichkeit, SSL-/TLS-geschützte Kommunikation zwischen Presto-Knoten durch Verschlüsselung der Daten während der Übertragung mithilfe einer Sicherheitskonfiguration zu aktivieren, wurde hinzugefügt. Weitere Informationen finden Sie unter [Verschlüsselung von Daten während der Übertragung](#).
- [Presto 7661](#) wurde rückportiert, wodurch der Anweisung EXPLAIN ANALYZE die Option VERBOSE hinzugefügt wurde, um detailliertere Low-Level-Statistiken über einen Abfrageplan bereitzustellen.

## 5.6.0 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>emr-ddb</code>	4.3.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.3.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.3.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-s3-dist-cp	2.5.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.17.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.2.1	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-2	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFS-Service auf Knotenebene zum Speichern von Blöcken.

Komponente	Version	Beschreibung
hadoop-hdfs-library	2.7.3-amzn-2	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httfs-server	2.7.3-amzn-2	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.7.3-amzn-2	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.7.3-amzn-2	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.

Komponente	Version	Beschreibung
hbase-hmaster	1.3.1	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.3.1	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.3.1	HBase-Befehlszeilen-Client.
hbase-rest-server	1.3.1	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.3.1	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.1.1-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.1.1-amzn-0	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.1.1-amzn-0	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.1.1-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.1.1-amzn-0	Hive-hbase client.



Komponente	Version	Beschreibung
hive-metastore-server	2.1.1-amzn-0	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.1.1-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.12.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.13.0	Bibliothek für Machine Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.9.0-HBase-1.2	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.9.0-HBase-1.2	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.

Komponente	Version	Beschreibung
presto-coordinator	0.170	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.170	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.16.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	2.1.1	Spark-Befehlszeilen-Clients.
spark-history-server	2.1.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.1.1	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.1.1	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.7.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.

Komponente	Version	Beschreibung
zookeeper-server	3.4.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.10	ZooKeeper Befehlszeilenclient.

## 5.6.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

### emr-5.6.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink <code>log4j-yarn-session.properties</code> .

Klassifizierungen	Beschreibung
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändern Sie die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.

Klassifizierungen	Beschreibung
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.

Klassifizierungen	Beschreibung
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie die Werte in der Anwendungsumgebung MapReduce .
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.

Klassifizierungen	Beschreibung
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-env	Ändern Sie die Werte in der presto-env.sh-Datei in Presto.
presto-node	Ändern Sie die Werte in der node.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.



Klassifizierungen	Beschreibung
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.5.4

### 5.5.4 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

### Informationen zur Anwendungsversion

	emr-5.5.4	emr-5.5.3	emr-5.5.2	emr-5.5.1
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	2.11,8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.2.0	1.2.0	1.2.0	1.2.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.0	1.3.0	1.3.0	1.3.0
HCatalog	2.1.1	2.1.1	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.1	2.1.1
Hudi	-	-	-	-
Hue	3.12.0	3.12.0	3.12.0	3.12.0
Iceberg	-	-	-	-

	emr-5.5.4	emr-5.5.3	emr-5.5.2	emr-5.5.1
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.9.0	4.9.0	4.9.0	4.9.0
Pig	0.16.0	0,16,0	0,16,0	0,16,0
Presto	0.170	0.170	0.170	0.170
Spark	2.1.0	2.1.0	2.1.0	2.1.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.1	0.7.1	0.7.1	0.7.1
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

## 5.5.4 Versionshinweise

Dies ist eine Patch-Version, um die AWS Signature Version 4-Authentifizierung für Anfragen an Amazon S3 hinzuzufügen. Alle Anwendungen und Komponenten sind dieselben wie in der vorherigen Amazon-EMR-Version.

### ⚠ Important

In dieser Release-Version verwendet Amazon EMR AWS Signature Version 4 ausschließlich zur Authentifizierung von Anfragen an Amazon S3. Weitere Informationen finden Sie unter [Was ist neu](#).

## 5.5.4 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>emr-ddb</code>	4.3.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.3.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.3.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-s3-dist-cp	2.5.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.16.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.2.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-2	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFS-Service auf Knotenebene zum Speichern von Blöcken.

Komponente	Version	Beschreibung
hadoop-hdfs-library	2.7.3-amzn-2	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httfs-server	2.7.3-amzn-2	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.7.3-amzn-2	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.7.3-amzn-2	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.

Komponente	Version	Beschreibung
hbase-hmaster	1.3.0	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.3.0	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.3.0	HBase-Befehlszeilen-Client.
hbase-rest-server	1.3.0	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.3.0	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.1.1-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.1.1-amzn-0	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.1.1-amzn-0	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.1.1-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.1.1-amzn-0	Hive-hbase client.

Komponente	Version	Beschreibung
hive-metastore-server	2.1.1-amzn-0	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.1.1-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.12.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.9.0-HBase-1.2	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.9.0-HBase-1.2	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.



Komponente	Version	Beschreibung
presto-coordinator	0.170	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.170	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.16.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	2.1.0	Spark-Befehlszeilen-Clients.
spark-history-server	2.1.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.1.0	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.1.0	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.7.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.

Komponente	Version	Beschreibung
zookeeper-server	3.4.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.10	ZooKeeper Befehlszeilenclient.

## 5.5.4 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

### emr-5.5.4-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink <code>log4j-yarn-session.properties</code> .

Klassifizierungen	Beschreibung
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.

Klassifizierungen	Beschreibung
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.

Klassifizierungen	Beschreibung
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie die Werte in der Anwendungsumgebung MapReduce .
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.

Klassifizierungen	Beschreibung
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.

Klassifizierungen	Beschreibung
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.5.3

### 5.5.3 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)



- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

## Informationen zur Anwendungsversion

	emr-5.5.3	emr-5.5.2	emr-5.5.1	emr-5.5.0
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	2.11,8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.2.0	1.2.0	1.2.0	1.2.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.0	1.3.0	1.3.0	1.3.0
HCatalog	2.1.1	2.1.1	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.1	2.1.1
Hudi	-	-	-	-
Hue	3.12.0	3.12.0	3.12.0	3.12.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-

	emr-5.5.3	emr-5.5.2	emr-5.5.1	emr-5.5.0
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.9.0	4.9.0	4.9.0	4.9.0
Pig	0.16.0	0,16,0	0,16,0	0,16,0
Presto	0.170	0.170	0.170	0.170
Spark	2.1.0	2.1.0	2.1.0	2.1.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.1	0.7.1	0.7.1	0.7.1
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

### 5.5.3 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.5.3. Änderungen beziehen sich auf Version 5.5.2.

Erste Version: 29. August 2018

Änderungen, Verbesserungen und behobene Probleme

- Diese Version behebt eine potenzielle Schwachstellen.

### 5.5.3 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>emr-ddb</code>	4.3.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.3.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.3.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-s3-dist-cp</code>	2.5.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
<code>emrfs</code>	2.16.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
flink-client	1.2.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-2	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-2	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.

Komponente	Version	Beschreibung
hadoop-httfs-server	2.7.3-amzn-2	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.7.3-amzn-2	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.7.3-amzn-2	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.3.0	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.3.0	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.

Komponente	Version	Beschreibung
hbase-client	1.3.0	HBase-Befehlszeilen-Client.
hbase-rest-server	1.3.0	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.3.0	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.1.1-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.1.1-amzn-0	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.1.1-amzn-0	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.1.1-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.1.1-amzn-0	Hive-hbase client.
hive-metastore-server	2.1.1-amzn-0	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.1.1-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.

Komponente	Version	Beschreibung
hue-server	3.12.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.9.0-HBase-1.2	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.9.0-HBase-1.2	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.170	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.170	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.16.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	2.1.0	Spark-Befehlszeilen-Clients.

Komponente	Version	Beschreibung
spark-history-server	2.1.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.1.0	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.1.0	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.7.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.10	ZooKeeper Befehlszeilenclient.



## 5.5.3 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

### emr-5.5.3-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Ändert die <code>log4j-cli.properties</code> -Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.

Klassifizierungen	Beschreibung
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.

Klassifizierungen	Beschreibung
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.

Klassifizierungen	Beschreibung
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie die Werte in der Anwendungsumgebung MapReduce .
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.

Klassifizierungen	Beschreibung
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.

Klassifizierungen	Beschreibung
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.5.2

### 5.5.2 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

### Informationen zur Anwendungsversion

	emr-5.5.2	emr-5.5.1	emr-5.5.0	emr-5.4.1
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt

	emr-5.5.2	emr-5.5.1	emr-5.5.0	emr-5.4.1
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.2.0	1.2.0	1.2.0	1.2.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.0	1.3.0	1.3.0	1.3.0
HCatalog	2.1.1	2.1.1	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.1	2.1.1
Hudi	-	-	-	-
Hue	3.12.0	3.12.0	3.12.0	3.11.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.9.0	4.9.0	4.9.0	4.9.0



	emr-5.5.2	emr-5.5.1	emr-5.5.0	emr-5.4.1
Pig	0.16.0	0,16,0	0,16,0	0,16,0
Presto	0.170	0.170	0.170	0.166
Spark	2.1.0	2.1.0	2.1.0	2.1.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.1	0.7.1	0.7.1	0.7.0
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.9

## 5.5.2 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.5.2. Änderungen beziehen sich auf Version 5.5.1.

Erste Version: 29. März 2018

Änderungen, Verbesserungen und behobene Probleme

- Der Amazon-Linux-Kernel des standardmäßigen Amazon-Linux-AMI für Amazon EMR wurde aktualisiert, um potenzielle Schwachstellen zu schließen.

## 5.5.2 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version

sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen myapp-component der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als 2.2-amzn-2 aufgeführt.

Komponente	Version	Beschreibung
emr-ddb	4.3.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.3.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.3.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.5.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.16.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.2.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem

Komponente	Version	Beschreibung
		zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-2	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-2	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httpfs-server	2.7.3-amzn-2	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.7.3-amzn-2	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider

Komponente	Version	Beschreibung
hadoop-mapred	2.7.3-amzn-2	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.3.0	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.3.0	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.3.0	HBase-Befehlszeilen-Client.
hbase-rest-server	1.3.0	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.3.0	Service, der einen Thrift-Endpunkt für HBase bereitstellt.

Komponente	Version	Beschreibung
hcatalog-client	2.1.1-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.1.1-amzn-0	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.1.1-amzn-0	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.1.1-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.1.1-amzn-0	Hive-hbase client.
hive-metastore-server	2.1.1-amzn-0	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.1.1-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.12.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.

Komponente	Version	Beschreibung
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.9.0-HBase-1.2	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.9.0-HBase-1.2	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.170	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.170	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.16.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	2.1.0	Spark-Befehlszeilen-Clients.
spark-history-server	2.1.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.1.0	In-Memory-Ausführungs-Engine für YARN.

Komponente	Version	Beschreibung
spark-yarn-slave	2.1.0	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.7.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.10	ZooKeeper Befehlszeilenclient.

## 5.5.2 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

## emr-5.5.2-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der capacity-scheduler.xml-Datei in Hadoop.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4 j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.



Klassifizierungen	Beschreibung
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.

Klassifizierungen	Beschreibung
hive-parquet-logging	Ändert die Werte in der <code>parquet-logging.properties</code> -Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von <code>hive-exec-log</code> Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von Hive. <code>llap-daemon-log</code>
hive-log4j2	Ändert die Werte in der <code>hive-log4j2.properties</code> -Datei in Hive.
hive-site	Ändert die Werte in der <code>hive-site.xml</code> -Datei in Hive.
hiveserver2-site	Ändert die Werte in der <code>hiveserver2-site.xml</code> -Datei von <code>Server2</code> in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der <code>httpfs-site.xml</code> -Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der <code>kms-acls.xml</code> -Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der <code>kms-log4j.properties</code> -Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der <code>kms-site.xml</code> -Datei in Hadoop.

Klassifizierungen	Beschreibung
mapred-env	Ändern Sie die Werte in der Anwendungsumgebung MapReduce .
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.

Klassifizierungen	Beschreibung
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.

Klassifizierungen	Beschreibung
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.5.1

### 5.5.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

#### Informationen zur Anwendungsversion

	emr-5.5.1	emr-5.5.0	emr-5.4.1	emr-5.4.0
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	2.11,8	2.11.8	2.11.8	2.11.8

	emr-5.5.1	emr-5.5.0	emr-5.4.1	emr-5.4.0
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.2.0	1.2.0	1.2.0	1.2.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.0	1.3.0	1.3.0	1.3.0
HCatalog	2.1.1	2.1.1	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.1	2.1.1
Hudi	-	-	-	-
Hue	3.12.0	3.12.0	3.11.0	3.11.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.9.0	4.9.0	4.9.0	4.9.0
Pig	0.16.0	0,16,0	0,16,0	0,16,0

	emr-5.5.1	emr-5.5.0	emr-5.4.1	emr-5.4.0
Presto	0.170	0.170	0.166	0.166
Spark	2.1.0	2.1.0	2.1.0	2.1.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.1	0.7.1	0.7.0	0.7.0
ZooKeeper	3.4.10	3.4.10	3.4.9	3.4.9

## 5.5.1 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.5.1. Änderungen beziehen sich auf Amazon-EMR-Version 5.5.0.

Erste Version: 22. Januar 2018

Änderungen, Verbesserungen und behobene Probleme

- Der Amazon-Linux-Kernel des standardmäßigen Amazon-Linux-AMI für Amazon EMR wurde aktualisiert, um Schwachstellen im Zusammenhang mit spekulativer Ausführung zu beheben (CVE-2017-5715, CVE-2017-5753 und CVE-2017-5754). Weitere Informationen finden Sie unter <https://aws.amazon.com/security/security-bulletins/AWS-2018-013/>.

## 5.5.1 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version



sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen myapp-component der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als 2.2-amzn-2 aufgeführt.

Komponente	Version	Beschreibung
emr-ddb	4.3.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.3.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.3.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.5.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.16.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.2.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem

Komponente	Version	Beschreibung
		zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-2	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-2	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httpfs-server	2.7.3-amzn-2	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.7.3-amzn-2	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider

Komponente	Version	Beschreibung
hadoop-mapred	2.7.3-amzn-2	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.3.0	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.3.0	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.3.0	HBase-Befehlszeilen-Client.
hbase-rest-server	1.3.0	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.3.0	Service, der einen Thrift-Endpunkt für HBase bereitstellt.

Komponente	Version	Beschreibung
hcatalog-client	2.1.1-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.1.1-amzn-0	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.1.1-amzn-0	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.1.1-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.1.1-amzn-0	Hive-hbase client.
hive-metastore-server	2.1.1-amzn-0	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.1.1-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.12.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.

Komponente	Version	Beschreibung
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.9.0-HBase-1.2	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.9.0-HBase-1.2	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.170	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.170	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.16.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	2.1.0	Spark-Befehlszeilen-Clients.
spark-history-server	2.1.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.1.0	In-Memory-Ausführungs-Engine für YARN.

Komponente	Version	Beschreibung
spark-yarn-slave	2.1.0	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.7.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.10	ZooKeeper Befehlszeilenclient.

### 5.5.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

## emr-5.5.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der capacity-scheduler.xml-Datei in Hadoop.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4 j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.

Klassifizierungen	Beschreibung
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.



Klassifizierungen	Beschreibung
hive-parquet-logging	Ändert die Werte in der <code>parquet-logging.properties</code> -Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von <code>hive-exec-log</code> Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von Hive. <code>llap-daemon-log</code>
hive-log4j2	Ändert die Werte in der <code>hive-log4j2.properties</code> -Datei in Hive.
hive-site	Ändert die Werte in der <code>hive-site.xml</code> -Datei in Hive.
hiveserver2-site	Ändert die Werte in der <code>hiveserver2-site.xml</code> -Datei von <code>Server2</code> in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der <code>httpfs-site.xml</code> -Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der <code>kms-acls.xml</code> -Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der <code>kms-log4j.properties</code> -Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der <code>kms-site.xml</code> -Datei in Hadoop.

Klassifizierungen	Beschreibung
mapred-env	Ändern Sie die Werte in der Anwendungsumgebung MapReduce .
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.

Klassifizierungen	Beschreibung
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.

Klassifizierungen	Beschreibung
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.5.0

### 5.5.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

### Informationen zur Anwendungsversion

	emr-5.5.0	emr-5.4.1	emr-5.4.0	emr-5.3.2
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	2.11,8	2.11.8	2.11.8	2.11.8

	emr-5.5.0	emr-5.4.1	emr-5.4.0	emr-5.3.2
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.2.0	1.2.0	1.2.0	1.1.4
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.0	1.3.0	1.3.0	1.2.3
HCatalog	2.1.1	2.1.1	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.1	2.1.1
Hudi	-	-	-	-
Hue	3.12.0	3.11.0	3.11.0	3.11.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.9.0	4.9.0	4.9.0	4.7.0
Pig	0.16.0	0,16,0	0,16,0	0,16,0

	emr-5.5.0	emr-5.4.1	emr-5.4.0	emr-5.3.2
Presto	0.170	0.166	0.166	0.157.1
Spark	2.1.0	2.1.0	2.1.0	2.1.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.1	0.7.0	0.7.0	0.6.2
ZooKeeper	3.4.10	3.4.9	3.4.9	3.4.9

## 5.5.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.5.0. Änderungen beziehen sich auf Amazon-EMR-Version 5.4.0.

Veröffentlichungsdatum: 26. April 2017

### Upgrades

- Hue 3.12
- Presto 0.170
- Zeppelin 0.7.1
- ZooKeeper 3.4.10

### Änderungen, Verbesserungen und behobene Probleme

- Spark
  - Spark-Patch ([SPARK-20115](#)) DAGScheduler wurde korrigiert, um alle verloren gegangenen Shuffle-Blöcke erneut zu berechnen, wenn der Shuffle-Service nicht verfügbar ist wurde auf Version 2.1.0 von Spark rückportiert, die in dieser Version enthalten ist.

- Flink
  - Flink ist nun in Scala 2.11 integriert. Wenn Sie die Scala-API und -Bibliotheken verwenden, empfehlen wir Scala 2.11 für Ihre Projekte.
  - Ein Problem wurde behoben, bei dem die Standardeinstellungen HADOOP\_CONF\_DIR und YARN\_CONF\_DIR nicht ordnungsgemäß festgelegt wurden, sodass `start-scala-shell.sh` nicht funktionierte. Außerdem wurde die Möglichkeit hinzugefügt, diese Werte mit `env.hadoop.conf.dir` und `env.yarn.conf.dir` in `/etc/flink/conf/flink-conf.yaml` oder der Konfigurationsklassifizierung `flink-conf` festzulegen.
  - Es wurde ein neuer EMR-spezifischer Befehl, `flink-scala-shell` als Wrapper für `start-scala-shell.sh` eingeführt. Wir empfehlen, diesen Befehl statt `start-scala-shell` zu verwenden. Der neue Befehl vereinfacht die Ausführung. Beispielsweise startet `flink-scala-shell -n 2` eine Flink Scala-Shell mit einer Aufgabenparallelität von 2.
  - Es wurde ein neuer EMR-spezifischer Befehl, `flink-yarn-session` als Wrapper für `yarn-session.sh` eingeführt. Wir empfehlen, diesen Befehl statt `yarn-session` zu verwenden. Der neue Befehl vereinfacht die Ausführung. Beispielsweise startet `flink-yarn-session -d -n 2` eine langlebige Flink-Sitzung in einem getrennten Status mit zwei Aufgabenmanagern.
  - Das Problem ([FLINK-6125](#)) [Commons httpclient is not shaded anymore in Flink 1.2](#) wurde behoben.
- Presto
  - Die LDAP-Authentifizierung wird jetzt unterstützt. Zur Verwendung von LDAP mit Presto in Amazon EMR; müssen Sie den HTTPS-Zugriff für den Presto-Koordinator (`http-server.https.enabled=true` in `config.properties`) aktivieren. Konfigurationsdetails finden Sie unter [LDAP Authentication](#) in der Presto-Dokumentation.
  - Unterstützung für `SHOW GRANTS` hinzugefügt.
- Amazon-EMR-Basis-Linux-AMI
  - Amazon-EMR-Versionen basieren jetzt auf Amazon Linux 2017.03. Weitere Informationen finden Sie in den [Amazon Linux AMI 2017.03-Versionshinweisen](#).
  - Python 2.6 wurde aus dem Amazon-EMR-Basis-Linux-Image entfernt. Python 2.7 und 3.4 sind standardmäßig installiert. Sie können Python 2.6 bei Bedarf manuell installieren.

## 5.5.0 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon



EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-`amzn-EmrVersion`. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>emr-ddb</code>	4.3.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.3.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.3.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-s3-dist-cp</code>	2.5.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
<code>emrfs</code>	2.16.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>flink-client</code>	1.2.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
<code>ganglia-monitor</code>	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus

Komponente	Version	Beschreibung
		dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-2	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-2	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httpfs-server	2.7.3-amzn-2	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.7.3-amzn-2	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider

Komponente	Version	Beschreibung
hadoop-mapred	2.7.3-amzn-2	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.3.0	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.3.0	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.3.0	HBase-Befehlszeilen-Client.
hbase-rest-server	1.3.0	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.3.0	Service, der einen Thrift-Endpunkt für HBase bereitstellt.

Komponente	Version	Beschreibung
hcatalog-client	2.1.1-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.1.1-amzn-0	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.1.1-amzn-0	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.1.1-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.1.1-amzn-0	Hive-hbase client.
hive-metastore-server	2.1.1-amzn-0	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.1.1-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.12.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.

Komponente	Version	Beschreibung
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.9.0-HBase-1.2	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.9.0-HBase-1.2	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.170	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.170	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.16.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	2.1.0	Spark-Befehlszeilen-Clients.
spark-history-server	2.1.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.1.0	In-Memory-Ausführungs-Engine für YARN.

Komponente	Version	Beschreibung
spark-yarn-slave	2.1.0	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.7.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.10	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.10	ZooKeeper Befehlszeilenclient.

### 5.5.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

## emr-5.5.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der capacity-scheduler.xml-Datei in Hadoop.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4 j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.

Klassifizierungen	Beschreibung
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.



Klassifizierungen	Beschreibung
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.

Klassifizierungen	Beschreibung
mapred-env	Ändern Sie die Werte in der Anwendungsumgebung MapReduce .
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.

Klassifizierungen	Beschreibung
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.

Klassifizierungen	Beschreibung
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.4.1

### 5.4.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

#### Informationen zur Anwendungsversion

	emr-5.4.1	emr-5.4.0	emr-5.3.2	emr-5.3.1
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	2.11,8	2.11.8	2.11.8	2.11.8

	emr-5.4.1	emr-5.4.0	emr-5.3.2	emr-5.3.1
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.2.0	1.2.0	1.1.4	1.1.4
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.0	1.3.0	1.2.3	1.2.3
HCatalog	2.1.1	2.1.1	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.1	2.1.1
Hudi	-	-	-	-
Hue	3.11.0	3.11.0	3.11.0	3.11.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.9.0	4.9.0	4.7.0	4.7.0
Pig	0.16.0	0,16,0	0,16,0	0,16,0

	emr-5.4.1	emr-5.4.0	emr-5.3.2	emr-5.3.1
Presto	0.166	0.166	0.157.1	0.157.1
Spark	2.1.0	2.1.0	2.1.0	2.1.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.0	0.7.0	0.6.2	0.6.2
ZooKeeper	3.4.9	3.4.9	3.4.9	3.4.9

### 5.4.1 Versionshinweise

Dies ist eine Patch-Version, um die AWS Signature Version 4-Authentifizierung für Anfragen an Amazon S3 hinzuzufügen. Alle Anwendungen und Komponenten sind dieselben wie in der vorherigen Amazon-EMR-Version.

#### Important

In dieser Release-Version verwendet Amazon EMR AWS Signature Version 4 ausschließlich zur Authentifizierung von Anfragen an Amazon S3. Weitere Informationen finden Sie unter [Was ist neu](#).

### 5.4.1 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version

sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen myapp-component der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als 2.2-amzn-2 aufgeführt.

Komponente	Version	Beschreibung
emr-ddb	4.2.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.3.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.2.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.4.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.15.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.2.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem



Komponente	Version	Beschreibung
		zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-1	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-1	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.7.3-amzn-1	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httpfs-server	2.7.3-amzn-1	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.7.3-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider

Komponente	Version	Beschreibung
hadoop-mapred	2.7.3-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-1	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.3-amzn-1	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-1	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.3.0	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.3.0	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.3.0	HBase-Befehlszeilen-Client.
hbase-rest-server	1.3.0	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.3.0	Service, der einen Thrift-Endpunkt für HBase bereitstellt.

Komponente	Version	Beschreibung
hcatalog-client	2.1.1-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.1.1-amzn-0	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.1.1-amzn-0	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.1.1-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.1.1-amzn-0	Hive-hbase client.
hive-metastore-server	2.1.1-amzn-0	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.1.1-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.11.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.

Komponente	Version	Beschreibung
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.9.0-HBase-1.2	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.9.0-HBase-1.2	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.166	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.166	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.16.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	2.1.0	Spark-Befehlszeilen-Clients.
spark-history-server	2.1.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.1.0	In-Memory-Ausführungs-Engine für YARN.

Komponente	Version	Beschreibung
spark-yarn-slave	2.1.0	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.7.0	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.9	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.9	ZooKeeper Befehlszeilenclient.

### 5.4.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

## emr-5.4.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der capacity-scheduler.xml-Datei in Hadoop.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4 j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.

Klassifizierungen	Beschreibung
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.

Klassifizierungen	Beschreibung
hive-parquet-logging	Ändert die Werte in der <code>parquet-logging.properties</code> -Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von <code>hive-exec-log</code> Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei <code>4j2.properties</code> von Hive. <code>llap-daemon-log</code>
hive-log4j2	Ändert die Werte in der <code>hive-log4j2.properties</code> -Datei in Hive.
hive-site	Ändert die Werte in der <code>hive-site.xml</code> -Datei in Hive.
hiveserver2-site	Ändert die Werte in der <code>hiveserver2-site.xml</code> -Datei von <code>Server2</code> in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der <code>httpfs-site.xml</code> -Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der <code>kms-acls.xml</code> -Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der <code>kms-log4j.properties</code> -Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der <code>kms-site.xml</code> -Datei in Hadoop.



Klassifizierungen	Beschreibung
mapred-env	Ändern Sie die Werte in der Anwendungsumgebung MapReduce .
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.

Klassifizierungen	Beschreibung
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.

Klassifizierungen	Beschreibung
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.4.0

### 5.4.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

### Informationen zur Anwendungsversion

	emr-5.4.0	emr-5.3.2	emr-5.3.1	emr-5.3.0
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	2.11,8	2.11.8	2.11.8	2.11.8

	emr-5.4.0	emr-5.3.2	emr-5.3.1	emr-5.3.0
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.2.0	1.1.4	1.1.4	1.1.4
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.0	1.2.3	1.2.3	1.2.3
HCatalog	2.1.1	2.1.1	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.1	2.1.1
Hudi	-	-	-	-
Hue	3.11.0	3.11.0	3.11.0	3.11.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.9.0	4.7.0	4.7.0	4.7.0
Pig	0.16.0	0,16,0	0,16,0	0,16,0

	emr-5.4.0	emr-5.3.2	emr-5.3.1	emr-5.3.0
Presto	0.166	0.157.1	0.157.1	0.157.1
Spark	2.1.0	2.1.0	2.1.0	2.1.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.0	0.6.2	0.6.2	0.6.2
ZooKeeper	3.4.9	3.4.9	3.4.9	3.4.9

## 5.4.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.4.0. Änderungen beziehen sich auf Amazon EMR Version 5.3.0.

Veröffentlichungsdatum: 8. März 2017

### Upgrades

- Upgrade auf Flink 1.2.0
- Upgrade auf HBase 1.3.0
- Upgrade auf Phoenix 4.9.0

#### Note

Wenn Sie von einer früheren Version von Amazon EMR auf Amazon EMR Version 5.4.0 oder höher upgraden und die sekundäre Indizierung verwenden, aktualisieren Sie lokale Indizes wie in der [Apache-Phoenix-Dokumentation](#) beschrieben. Amazon EMR entfernt die erforderlichen Konfigurationen aus der `hbase-site`-Klassifizierung, die Indizes müssen jedoch neu gefüllt werden. Online- und Offline-Upgrades von Indizes

werden unterstützt. Online-Upgrades sind die Standardeinstellung. Dies bedeutet, dass Indizes neu gefüllt werden, während die Initialisierung von Phoenix-Clients Version 4.8.0 oder höher ausgeführt wird. Um Offline-Upgrades festzulegen, geben Sie für die `phoenix.client.localIndexUpgrade`-Konfiguration in der `phoenix-site`-Klassifizierung `false` und dann SSH für den Primärknoten an, um `psql [zookeeper] -1` auszuführen.

- Upgrade auf Presto 0.166
- Upgrade auf Zeppelin 0.7.0

### Änderungen und Verbesserungen

- Es werden jetzt r4-Instances unterstützt. Sehen Sie unter [Amazon-EC2-Instance-Typen](#).

## 5.4.0 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>emr-ddb</code>	4.2.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-goodies	2.3.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.2.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.4.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.15.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.2.0	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.



Komponente	Version	Beschreibung
hadoop-client	2.7.3-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-1	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-1	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.7.3-amzn-1	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httpfs-server	2.7.3-amzn-1	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.7.3-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.7.3-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-1	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.3-amzn-1	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.

Komponente	Version	Beschreibung
hadoop-yarn-timeline-server	2.7.3-amzn-1	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.3.0	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.3.0	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.3.0	HBase-Befehlszeilen-Client.
hbase-rest-server	1.3.0	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.3.0	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.1.1-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.1.1-amzn-0	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.1.1-amzn-0	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.

Komponente	Version	Beschreibung
hive-client	2.1.1-amzn-0	Hive-Befehlszeilen-Client.
hive-hbase	2.1.1-amzn-0	Hive-hbase client.
hive-metastore-server	2.1.1-amzn-0	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server2	2.1.1-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.11.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.9.0-HBase-1.2	Die Phoenix-Bibliotheken für den Server und den Client

Komponente	Version	Beschreibung
phoenix-query-server	4.9.0-HBase-1.2	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.166	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.166	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.16.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	2.1.0	Spark-Befehlszeilen-Clients.
spark-history-server	2.1.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.1.0	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.1.0	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.

Komponente	Version	Beschreibung
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.7.0	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.9	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.9	ZooKeeper Befehlszeilenclient.

## 5.4.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

### emr-5.4.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.

Klassifizierungen	Beschreibung
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4 j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.

Klassifizierungen	Beschreibung
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.

Klassifizierungen	Beschreibung
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie die Werte in der Anwendungsumgebung MapReduce .
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.



Klassifizierungen	Beschreibung
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.

Klassifizierungen	Beschreibung
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.3.2

### 5.3.2 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

### Informationen zur Anwendungsversion

	emr-5.3.2	emr-5.3.1	emr-5.3.0	emr-5.2.3
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	2.11,8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.1.4	1.1.4	1.1.4	1.1.3
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.3	1.2.3	1.2.3	1.2.3
HCatalog	2.1.1	2.1.1	2.1.1	2.1.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.1	2.1.0
Hudi	-	-	-	-
Hue	3.11.0	3.11.0	3.11.0	3.10.0
Iceberg	-	-	-	-

	emr-5.3.2	emr-5.3.1	emr-5.3.0	emr-5.2.3
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.3.0	4.3.0	4.3.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.16.0	0,16,0	0,16,0	0,16,0
Presto	0.157.1	0.157.1	0.157.1	0.157.1
Spark	2.1.0	2.1.0	2.1.0	2.0.2
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.6.2	0.6.2	0.6.2	0.6.2
ZooKeeper	3.4.9	3.4.9	3.4.9	3.4.9

### 5.3.2 Versionshinweise

Dies ist eine Patch-Version, um die AWS Signature Version 4-Authentifizierung für Anfragen an Amazon S3 hinzuzufügen. Alle Anwendungen und Komponenten sind dieselben wie in der vorherigen Amazon-EMR-Version.

**⚠ Important**

In dieser Release-Version verwendet Amazon EMR AWS Signature Version 4 ausschließlich zur Authentifizierung von Anfragen an Amazon S3. Weitere Informationen finden Sie unter [Was ist neu](#).

### 5.3.2 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>emr-ddb</code>	4.2.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.2.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.2.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-s3-dist-cp	2.4.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.14.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.1.4	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-1	HDFS-Service auf Knotenebene zum Speichern von Blöcken.

Komponente	Version	Beschreibung
hadoop-hdfs-library	2.7.3-amzn-1	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.7.3-amzn-1	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httpfs-server	2.7.3-amzn-1	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.7.3-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.7.3-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-1	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.3-amzn-1	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-1	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.



Komponente	Version	Beschreibung
hbase-hmaster	1.2.3	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.2.3	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.2.3	HBase-Befehlszeilen-Client.
hbase-rest-server	1.2.3	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.2.3	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.1.1-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.1.1-amzn-0	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.1.1-amzn-0	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.1.1-amzn-0	Hive-Befehlszeilen-Client.

Komponente	Version	Beschreibung
hive-metastore-server	2.1.1-amzn-0	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server	2.1.1-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.11.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.7.0-HBase-1.2	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.7.0-HBase-1.2	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.

Komponente	Version	Beschreibung
presto-coordinator	0.157.1	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.157.1	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.16.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	2.1.0	Spark-Befehlszeilen-Clients.
spark-history-server	2.1.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.1.0	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.1.0	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.6.2	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.

Komponente	Version	Beschreibung
zookeeper-server	3.4.9	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.9	ZooKeeper Befehlszeilenclient.

### 5.3.2 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

#### emr-5.3.2-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink <code>log4j-yarn-session.properties</code> .

Klassifizierungen	Beschreibung
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.

Klassifizierungen	Beschreibung
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.

Klassifizierungen	Beschreibung
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie die Werte in der Anwendungsumgebung MapReduce .
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.

Klassifizierungen	Beschreibung
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.



Klassifizierungen	Beschreibung
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.

Klassifizierungen	Beschreibung
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.3.1

### 5.3.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)

- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

## Informationen zur Anwendungsversion

	emr-5.3.1	emr-5.3.0	emr-5.2.3	emr-5.2.2
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	2.11,8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.1.4	1.1.4	1.1.3	1.1.3
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.3	1.2.3	1.2.3	1.2.3
HCatalog	2.1.1	2.1.1	2.1.0	2.1.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.0	2.1.0
Hudi	-	-	-	-
Hue	3.11.0	3.11.0	3.10.0	3.10.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-

	emr-5.3.1	emr-5.3.0	emr-5.2.3	emr-5.2.2
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.3.0	4.3.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.16.0	0,16,0	0,16,0	0,16,0
Presto	0.157.1	0.157.1	0.157.1	0.157.1
Spark	2.1.0	2.1.0	2.0.2	2.0.2
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.6.2	0.6.2	0.6.2	0.6.2
ZooKeeper	3.4.9	3.4.9	3.4.9	3.4.9

### 5.3.1 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon EMR Version 5.3.1. Änderungen beziehen sich auf Amazon EMR Version 5.3.0.

Veröffentlichungsdatum: 7. Februar 2017

Kleinere Änderungen am Backport von Zeppelin-Patches und an der Aktualisierung des Standard-AMI für Amazon EMR.

### 5.3.1 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>emr-ddb</code>	4.2.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.2.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.2.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-s3-dist-cp</code>	2.4.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
<code>emrfs</code>	2.14.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
flink-client	1.1.4	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-1	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-1	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.7.3-amzn-1	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.

Komponente	Version	Beschreibung
hadoop-httfs-server	2.7.3-amzn-1	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.7.3-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.7.3-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-1	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.3-amzn-1	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-1	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.2.3	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.2.3	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.

Komponente	Version	Beschreibung
hbase-client	1.2.3	HBase-Befehlszeilen-Client.
hbase-rest-server	1.2.3	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.2.3	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.1.1-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.1.1-amzn-0	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.1.1-amzn-0	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.1.1-amzn-0	Hive-Befehlszeilen-Client.
hive-metastore-server	2.1.1-amzn-0	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server	2.1.1-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.



Komponente	Version	Beschreibung
hue-server	3.11.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.7.0-HBase-1.2	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.7.0-HBase-1.2	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.157.1	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.157.1	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.16.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	2.1.0	Spark-Befehlszeilen-Clients.

Komponente	Version	Beschreibung
spark-history-server	2.1.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.1.0	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.1.0	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.6.2	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.9	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.9	ZooKeeper Befehlszeilenclient.

### 5.3.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

#### emr-5.3.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Ändert die <code>log4j-cli.properties</code> -Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.

Klassifizierungen	Beschreibung
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.

Klassifizierungen	Beschreibung
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.

Klassifizierungen	Beschreibung
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie die Werte in der Anwendungsumgebung MapReduce .
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.

Klassifizierungen	Beschreibung
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.



Klassifizierungen	Beschreibung
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.3.0

### 5.3.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

### Informationen zur Anwendungsversion

	emr-5.3.0	emr-5.2.3	emr-5.2.2	emr-5.2.1
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt

	emr-5.3.0	emr-5.2.3	emr-5.2.2	emr-5.2.1
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.1.4	1.1.3	1.1.3	1.1.3
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.3	1.2.3	1.2.3	1.2.3
HCatalog	2.1.1	2.1.0	2.1.0	2.1.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.0	2.1.0	2.1.0
Hudi	-	-	-	-
Hue	3.11.0	3.10.0	3.10.0	3.10.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.3.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0

	emr-5.3.0	emr-5.2.3	emr-5.2.2	emr-5.2.1
Pig	0.16.0	0,16,0	0,16,0	0,16,0
Presto	0.157.1	0.157.1	0.157.1	0.157.1
Spark	2.1.0	2.0.2	2.0.2	2.0.2
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.6.2	0.6.2	0.6.2	0.6.2
ZooKeeper	3.4.9	3.4.9	3.4.9	3.4.9

### 5.3.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon EMR Version 5.3.0. Änderungen beziehen sich auf Amazon EMR Version 5.2.1.

Veröffentlichungsdatum: 26. Januar 2017

#### Upgrades

- Upgrade auf Hive 2.1.1
- Upgrade auf Hue 3.11.0
- Upgrade auf Spark 2.1.0
- Upgrade auf Oozie 4.3.0
- Upgrade auf Flink 1.1.4

## Änderungen und Verbesserungen

- Ein Patch für Hue wurde hinzugefügt, mit dem Sie die Einstellung `interpreters_shown_on_wheel` verwenden können, um zu konfigurieren, was Interpreter in der Notebook-Auswahl als Erstes anzeigen, unabhängig von ihrer Reihenfolge in der Datei `hue.ini`.
- Die Konfigurationsklassifizierung `hive-parquet-logging` wurde hinzugefügt, mit der Sie Werte in der Hive-Datei `parquet-logging.properties` konfigurieren können.

### 5.3.0 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-*amzn*-*EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>emr-ddb</code>	4.2.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.2.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.2.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-s3-dist-cp	2.4.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.14.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.1.4	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-1	HDFS-Service auf Knotenebene zum Speichern von Blöcken.

Komponente	Version	Beschreibung
hadoop-hdfs-library	2.7.3-amzn-1	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.7.3-amzn-1	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httfs-server	2.7.3-amzn-1	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.7.3-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.7.3-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-1	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.3-amzn-1	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-1	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.

Komponente	Version	Beschreibung
hbase-hmaster	1.2.3	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.2.3	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.2.3	HBase-Befehlszeilen-Client.
hbase-rest-server	1.2.3	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.2.3	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.1.1-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.1.1-amzn-0	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.1.1-amzn-0	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.1.1-amzn-0	Hive-Befehlszeilen-Client.

Komponente	Version	Beschreibung
hive-metastore-server	2.1.1-amzn-0	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server	2.1.1-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.11.0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.52	MySQL-Datenbankserver.
oozie-client	4.3.0	Oozie-Befehlszeilen-Client.
oozie-server	4.3.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.7.0-HBase-1.2	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.7.0-HBase-1.2	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.



Komponente	Version	Beschreibung
presto-coordinator	0.157.1	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.157.1	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.16.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	2.1.0	Spark-Befehlszeilen-Clients.
spark-history-server	2.1.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.1.0	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.1.0	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.6.2	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.

Komponente	Version	Beschreibung
zookeeper-server	3.4.9	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.9	ZooKeeper Befehlszeilenclient.

### 5.3.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

#### emr-5.3.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink <code>log4j-yarn-session.properties</code> .

Klassifizierungen	Beschreibung
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.

Klassifizierungen	Beschreibung
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-parquet-logging	Ändert die Werte in der parquet-logging.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.

Klassifizierungen	Beschreibung
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie die Werte in der Anwendungsumgebung MapReduce .
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.

Klassifizierungen	Beschreibung
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.

Klassifizierungen	Beschreibung
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.2.3

### 5.2.3 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)



- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

## Informationen zur Anwendungsversion

	emr-5.2.3	emr-5.2.2	emr-5.2.1	emr-5.2.0
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	2.11,8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.1.3	1.1.3	1.1.3	1.1.3
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.3	1.2.3	1.2.3	1.2.3
HCatalog	2.1.0	2.1.0	2.1.0	2.1.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.0	2.1.0	2.1.0	2.1.0
Hudi	-	-	-	-
Hue	3.10.0	3.10.0	3.10.0	3.10.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-

	emr-5.2.3	emr-5.2.2	emr-5.2.1	emr-5.2.0
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.16.0	0,16,0	0,16,0	0,16,0
Presto	0.157.1	0.157.1	0.157.1	0.152.3
Spark	2.0.2	2.0.2	2.0.2	2.0.2
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.6.2	0.6.2	0.6.2	0.6.2
ZooKeeper	3.4.9	3.4.9	3.4.9	3.4.8

### 5.2.3 Versionshinweise

Dies ist eine Patch-Version, um die AWS Signature Version 4-Authentifizierung für Anfragen an Amazon S3 hinzuzufügen. Alle Anwendungen und Komponenten sind dieselben wie in der vorherigen Amazon-EMR-Version.

### ⚠ Important

In dieser Release-Version verwendet Amazon EMR AWS Signature Version 4 ausschließlich zur Authentifizierung von Anfragen an Amazon S3. Weitere Informationen finden Sie unter [Was ist neu](#).

## 5.2.3 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>emr-ddb</code>	4.2.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.2.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.2.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-s3-dist-cp	2.4.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.13.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.1.3	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-1	HDFS-Service auf Knotenebene zum Speichern von Blöcken.

Komponente	Version	Beschreibung
hadoop-hdfs-library	2.7.3-amzn-1	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.7.3-amzn-1	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httpfs-server	2.7.3-amzn-1	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.7.3-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.7.3-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-1	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.3-amzn-1	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-1	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.

Komponente	Version	Beschreibung
hbase-hmaster	1.2.3	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.2.3	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.2.3	HBase-Befehlszeilen-Client.
hbase-rest-server	1.2.3	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.2.3	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.1.0-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.1.0-amzn-0	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.1.0-amzn-0	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.1.0-amzn-0	Hive-Befehlszeilen-Client.

Komponente	Version	Beschreibung
hive-metastore-server	2.1.0-amzn-0	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server	2.1.0-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.10.0-amzn-0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.52	MySQL-Datenbankserver.
oozie-client	4.2.0	Oozie-Befehlszeilen-Client.
oozie-server	4.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.7.0-HBase-1.2	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.7.0-HBase-1.2	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.

Komponente	Version	Beschreibung
presto-coordinator	0.157.1	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.157.1	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.16.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	2.0.2	Spark-Befehlszeilen-Clients.
spark-history-server	2.0.2	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.0.2	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.0.2	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.23	Apache HTTP-Server.
zeppelin-server	0.6.2	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.



Komponente	Version	Beschreibung
zookeeper-server	3.4.9	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.9	ZooKeeper Befehlszeilenclient.

### 5.2.3 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

#### emr-5.2.3-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink <code>log4j-yarn-session.properties</code> .

Klassifizierungen	Beschreibung
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.

Klassifizierungen	Beschreibung
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.

Klassifizierungen	Beschreibung
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie die Werte in der Anwendungsumgebung MapReduce .
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.

Klassifizierungen	Beschreibung
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.

Klassifizierungen	Beschreibung
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.2.2

### 5.2.2 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

## Informationen zur Anwendungsversion

	emr-5.2.2	emr-5.2.1	emr-5.2.0	emr-5.1.1
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.1.3	1.1.3	1.1.3	1.1.3
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.3	1.2.3	1.2.3	1.2.3
HCatalog	2.1.0	2.1.0	2.1.0	2.1.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.0	2.1.0	2.1.0	2.1.0
Hudi	-	-	-	-
Hue	3.10.0	3.10.0	3.10.0	3.10.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-



	emr-5.2.2	emr-5.2.1	emr-5.2.0	emr-5.1.1
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.16.0	0,16,0	0,16,0	0,16,0
Presto	0.157.1	0.157.1	0.152.3	0.152.3
Spark	2.0.2	2.0.2	2.0.2	2.0.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.6.2	0.6.2	0.6.2	0.6.2
ZooKeeper	3.4.9	3.4.9	3.4.8	3.4.8

## 5.2.2 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon EMR Version 5.2.2. Änderungen beziehen sich auf Amazon EMR Version 5.2.1.

Veröffentlichungsdatum: 2. Mai 2017

Bekannte Probleme aus den früheren Versionen, die behoben wurden

- [SPARK-194459](#) wurde rückportiert. Damit wurde ein Problem behoben, bei dem beim Lesen aus einer ORC-Tabelle mit CHAR/VARCHAR-Spalten Fehler auftreten konnten.

## 5.2.2 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>emr-ddb</code>	4.2.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.2.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.2.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-s3-dist-cp</code>	2.4.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
<code>emrfs</code>	2.13.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
flink-client	1.1.3	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-1	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-1	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.7.3-amzn-1	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.

Komponente	Version	Beschreibung
hadoop-httfs-server	2.7.3-amzn-1	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.7.3-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.7.3-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-1	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.3-amzn-1	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-1	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.2.3	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.2.3	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.

Komponente	Version	Beschreibung
hbase-client	1.2.3	HBase-Befehlszeilen-Client.
hbase-rest-server	1.2.3	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.2.3	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.1.0-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.1.0-amzn-0	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.1.0-amzn-0	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.1.0-amzn-0	Hive-Befehlszeilen-Client.
hive-metastore-server	2.1.0-amzn-0	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server	2.1.0-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.

Komponente	Version	Beschreibung
hue-server	3.10.0-amzn-0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.52	MySQL-Datenbankserver.
oozie-client	4.2.0	Oozie-Befehlszeilen-Client.
oozie-server	4.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.7.0-HBase-1.2	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.7.0-HBase-1.2	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.157.1	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.157.1	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.16.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	2.0.2	Spark-Befehlszeilen-Clients.

Komponente	Version	Beschreibung
spark-history-server	2.0.2	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.0.2	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.0.2	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.23	Apache HTTP-Server.
zeppelin-server	0.6.2	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.9	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.9	ZooKeeper Befehlszeilenclient.

## 5.2.2 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

### emr-5.2.2-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Ändert die <code>log4j-cli.properties</code> -Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.



Klassifizierungen	Beschreibung
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.

Klassifizierungen	Beschreibung
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.

Klassifizierungen	Beschreibung
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie die Werte in der Anwendungsumgebung MapReduce .
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.

Klassifizierungen	Beschreibung
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.

Klassifizierungen	Beschreibung
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.2.1

### 5.2.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

#### Informationen zur Anwendungsversion

	emr-5.2.1	emr-5.2.0	emr-5.1.1	emr-5.1.0
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	2.11,8	2.11.8	2.11.8	2.11.8

	emr-5.2.1	emr-5.2.0	emr-5.1.1	emr-5.1.0
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.1.3	1.1.3	1.1.3	1.1.3
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.3	1.2.3	1.2.3	1.2.3
HCatalog	2.1.0	2.1.0	2.1.0	2.1.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.0	2.1.0	2.1.0	2.1.0
Hudi	-	-	-	-
Hue	3.10.0	3.10.0	3.10.0	3.10.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.16.0	0,16,0	0,16,0	0,16,0

	emr-5.2.1	emr-5.2.0	emr-5.1.1	emr-5.1.0
Presto	0.157.1	0.152.3	0.152.3	0.152.3
Spark	2.0.2	2.0.2	2.0.1	2.0.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.6.2	0.6.2	0.6.2	0.6.2
ZooKeeper	3.4.9	3.4.8	3.4.8	3.4.8

## 5.2.1 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon EMR Version 5.2.1. Änderungen beziehen sich auf Amazon EMR Version 5.2.0.

Veröffentlichungsdatum: 29. Dezember 2016

### Upgrades

- Upgrade auf Presto 0.157.1. Weitere Informationen finden Sie in den [Presto-Versionshinweisen](#) in der Presto-Dokumentation.
- Upgrade auf ZooKeeper 3.4.9. Weitere Informationen finden Sie in den [ZooKeeper Versionshinweisen](#) in der ZooKeeper Apache-Dokumentation.

### Änderungen und Verbesserungen

- Unterstützung hinzugefügt für den Amazon-EC2-Instanz-Typ m4.16xlarge in Amazon-EMR-Version 4.8.3 und höher, ausgenommen 5.0.0, 5.0.3 und 5.2.0.
- Amazon-EMR-Versionen basieren jetzt auf Amazon Linux 2016.09. Weitere Informationen finden Sie unter <https://aws.amazon.com/amazon-linux-ami/2016.09-release-notes/>.



- Der Speicherort von Flink- und YARN-Konfigurationspfaden wird nun standardmäßig in `/etc/default/flink` festgelegt. Sie brauchen die Umgebungsvariablen `FLINK_CONF_DIR` und `HADOOP_CONF_DIR` bei der Ausführung des Treiberskripts `flink` oder `yarn-session.sh` zum Starten von Flink-Aufträgen nicht festlegen.
- Unterstützung für `FlinkKinesisConsumer` Klassen hinzugefügt.

Bekannte Probleme aus den früheren Versionen, die behoben wurden

- Es wurde ein Problem in Hadoop behoben, bei dem der `ReplicationMonitor Thread` aufgrund eines Wettlaufs zwischen Replikation und Löschung derselben Datei in einem großen Cluster für lange Zeit hängen bleiben konnte.
- Es wurde ein Problem behoben, bei dem `ControlledJob #toString` mit einer Nullzeiger-Ausnahme (NPE) fehlschlug, wenn der Jobstatus nicht erfolgreich aktualisiert wurde.

## 5.2.1 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>emr-ddb</code>	4.2.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-goodies	2.2.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.2.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.4.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.13.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.1.3	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.

Komponente	Version	Beschreibung
hadoop-client	2.7.3-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-1	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-1	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.7.3-amzn-1	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httpfs-server	2.7.3-amzn-1	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.7.3-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.7.3-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-1	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.3-amzn-1	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.

Komponente	Version	Beschreibung
hadoop-yarn-timeline-server	2.7.3-amzn-1	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.2.3	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.2.3	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.2.3	HBase-Befehlszeilen-Client.
hbase-rest-server	1.2.3	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.2.3	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.1.0-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.1.0-amzn-0	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.1.0-amzn-0	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.

Komponente	Version	Beschreibung
hive-client	2.1.0-amzn-0	Hive-Befehlszeilen-Client.
hive-metastore-server	2.1.0-amzn-0	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server	2.1.0-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.10.0-amzn-0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.52	MySQL-Datenbankserver.
oozie-client	4.2.0	Oozie-Befehlszeilen-Client.
oozie-server	4.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.7.0-HBase-1.2	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.7.0-HBase-1.2	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.

Komponente	Version	Beschreibung
presto-coordinator	0.157.1	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.157.1	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.16.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	2.0.2	Spark-Befehlszeilen-Clients.
spark-history-server	2.0.2	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.0.2	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.0.2	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.23	Apache HTTP-Server.
zeppelin-server	0.6.2	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.

Komponente	Version	Beschreibung
zookeeper-server	3.4.9	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.9	ZooKeeper Befehlszeilenclient.

## 5.2.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

### emr-5.2.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink <code>log4j-yarn-session.properties</code> .

Klassifizierungen	Beschreibung
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.



Klassifizierungen	Beschreibung
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.

Klassifizierungen	Beschreibung
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie die Werte in der Anwendungsumgebung MapReduce .
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.

Klassifizierungen	Beschreibung
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.

Klassifizierungen	Beschreibung
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.2.0

### 5.2.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

## Informationen zur Anwendungsversion

	emr-5.2.0	emr-5.1.1	emr-5.1.0	emr-5.0.3
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.1.3	1.1.3	1.1.3	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.3	1.2.3	1.2.3	1.2.2
HCatalog	2.1.0	2.1.0	2.1.0	2.1.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.0	2.1.0	2.1.0	2.1.0
Hudi	-	-	-	-
Hue	3.10.0	3.10.0	3.10.0	3.10.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-

	emr-5.2.0	emr-5.1.1	emr-5.1.0	emr-5.0.3
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.16.0	0,16,0	0,16,0	0,16,0
Presto	0.152.3	0.152.3	0.152.3	0.152.3
Spark	2.0.2	2.0.1	2.0.1	2.0.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.6.2	0.6.2	0.6.2	0.6.1
ZooKeeper	3.4.8	3.4.8	3.4.8	3.4.8

## 5.2.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon EMR Version 5.2.0. Änderungen beziehen sich auf Amazon EMR Version 5.1.0.

Veröffentlichungsdatum: 21. November 2016

### Änderungen und Verbesserungen

- Der Amazon-S3-Speichermodus für HBase wurde hinzugefügt.
- Jetzt können Sie einen Amazon-S3-Speicherort für das HBase-Stammverzeichnis angeben. Weitere Informationen finden Sie unter [HBase in Amazon S3](#).

## Upgrades

- Upgrade auf Spark 2.0.2

### Bekannte Probleme aus den früheren Versionen, die behoben wurden

- Es wurde ein Problem mit der Beschränkung von /mnt auf 2 TB auf ausschließlichen EBS-Instance-Typen behoben.
- Es wurde ein Problem mit Instance-Controller und Logpusher-Protokollen behoben, die an die entsprechenden OUT-Dateien anstelle ihrer üblichen log4j-konfigurierten LOG-Dateien, die stündlich rotieren, ausgegeben wurden. Da die OUT-Dateien nicht rotieren, ist letztendlich die /emr-Partition belegt. Dieses Problem wirkt sich nur auf Hardware Virtual Machine (HVM)-Instance-Typen aus.

## 5.2.0 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
emr-ddb	4.1.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.



Komponente	Version	Beschreibung
emr-goodies	2.1.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.2.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.4.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.12.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
flink-client	1.1.3	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.

Komponente	Version	Beschreibung
hadoop-client	2.7.3-amzn-0	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-0	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-0	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.7.3-amzn-0	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httpfs-server	2.7.3-amzn-0	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.7.3-amzn-0	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.7.3-amzn-0	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-0	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.3-amzn-0	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.

Komponente	Version	Beschreibung
hadoop-yarn-timeline-server	2.7.3-amzn-0	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.2.3	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.2.3	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.2.3	HBase-Befehlszeilen-Client.
hbase-rest-server	1.2.3	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.2.3	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.1.0-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.1.0-amzn-0	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.1.0-amzn-0	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.

Komponente	Version	Beschreibung
hive-client	2.1.0-amzn-0	Hive-Befehlszeilen-Client.
hive-metastore-server	2.1.0-amzn-0	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server	2.1.0-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.10.0-amzn-0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.52	MySQL-Datenbankserver.
oozie-client	4.2.0	Oozie-Befehlszeilen-Client.
oozie-server	4.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.7.0-HBase-1.2	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.7.0-HBase-1.2	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.

Komponente	Version	Beschreibung
presto-coordinator	0.152.3	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.152.3	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.16.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	2.0.2	Spark-Befehlszeilen-Clients.
spark-history-server	2.0.2	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.0.2	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.0.2	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.23	Apache HTTP-Server.
zeppelin-server	0.6.2	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.

Komponente	Version	Beschreibung
zookeeper-server	3.4.8	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.8	ZooKeeper Befehlszeilenclient.

## 5.2.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

### emr-5.2.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink <code>log4j-yarn-session.properties</code> .

Klassifizierungen	Beschreibung
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache HBase.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.

Klassifizierungen	Beschreibung
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.



Klassifizierungen	Beschreibung
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie die Werte in der Anwendungsumgebung MapReduce .
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.

Klassifizierungen	Beschreibung
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.

Klassifizierungen	Beschreibung
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.1.1

### 5.1.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

## Informationen zur Anwendungsversion

	emr-5.1.1	emr-5.1.0	emr-5.0.3	emr-5.0.2
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.1.3	1.1.3	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.3	1.2.3	1.2.2	1.2.2
HCatalog	2.1.0	2.1.0	2.1.0	2.1.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.2
Hive	2.1.0	2.1.0	2.1.0	2.1.0
Hudi	-	-	-	-
Hue	3.10.0	3.10.0	3.10.0	3.10.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-

	emr-5.1.1	emr-5.1.0	emr-5.0.3	emr-5.0.2
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.16.0	0,16,0	0,16,0	0,16,0
Presto	0.152.3	0.152.3	0.152.3	0.150
Spark	2.0.1	2.0.1	2.0.1	2.0.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.6.2	0.6.2	0.6.1	0.6.1
ZooKeeper	3.4.8	3.4.8	3.4.8	3.4.8

### 5.1.1 Versionshinweise

Dies ist eine Patch-Version, um die AWS Signature Version 4-Authentifizierung für Anfragen an Amazon S3 hinzuzufügen. Alle Anwendungen und Komponenten sind dieselben wie in der vorherigen Amazon-EMR-Version.

#### Important

In dieser Release-Version verwendet Amazon EMR AWS Signature Version 4 ausschließlich zur Authentifizierung von Anfragen an Amazon S3. Weitere Informationen finden Sie unter [Was ist neu](#).

## 5.1.1 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>emr-ddb</code>	4.1.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.1.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.2.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-s3-dist-cp</code>	2.4.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
<code>emrfs</code>	2.11.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
flink-client	1.1.3	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-0	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-0	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-0	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.7.3-amzn-0	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.



Komponente	Version	Beschreibung
hadoop-httfs-server	2.7.3-amzn-0	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.7.3-amzn-0	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.7.3-amzn-0	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-0	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.3-amzn-0	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-0	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.2.3	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.2.3	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.

Komponente	Version	Beschreibung
hbase-client	1.2.3	HBase-Befehlszeilen-Client.
hbase-rest-server	1.2.3	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.2.3	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.1.0-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.1.0-amzn-0	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.1.0-amzn-0	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.1.0-amzn-0	Hive-Befehlszeilen-Client.
hive-metastore-server	2.1.0-amzn-0	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server	2.1.0-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.

Komponente	Version	Beschreibung
hue-server	3.10.0-amzn-0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.52	MySQL-Datenbankserver.
oozie-client	4.2.0	Oozie-Befehlszeilen-Client.
oozie-server	4.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.7.0-HBase-1.2	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.7.0-HBase-1.2	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.152.3	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.152.3	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.16.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	2.0.1	Spark-Befehlszeilen-Clients.

Komponente	Version	Beschreibung
spark-history-server	2.0.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.0.1	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.0.1	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.23	Apache HTTP-Server.
zeppelin-server	0.6.2	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.8	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.8	ZooKeeper Befehlszeilenclient.

## 5.1.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

### emr-5.1.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die <code>flink-conf.yaml</code> -Einstellungen.
flink-log4j	Ändert die <code>log4j.properties</code> -Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Ändert die <code>log4j-cli.properties</code> -Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase-env	Ändert die Werte in der HBase-Umgebung.

Klassifizierungen	Beschreibung
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.

Klassifizierungen	Beschreibung
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie die Werte in der Anwendungsumgebung MapReduce .

Klassifizierungen	Beschreibung
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.



Klassifizierungen	Beschreibung
presto-connector-cassandra	Ändert die Werte in der <code>cassandra.properties</code> -Datei in Presto.
presto-connector-hive	Ändert die Werte in der <code>hive.properties</code> -Datei in Presto.
presto-connector-jmx	Ändert die Werte in der <code>jmx.properties</code> -Datei in Presto.
presto-connector-kafka	Ändert die Werte in der <code>kafka.properties</code> -Datei in Presto.
presto-connector-localfile	Ändert die Werte in der <code>localfile.properties</code> -Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der <code>mongodb.properties</code> -Datei in Presto.
presto-connector-mysql	Ändert die Werte in der <code>mysql.properties</code> -Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der <code>postgresql.properties</code> -Datei in Presto.
presto-connector-raptor	Ändert die Werte in der <code>raptor.properties</code> -Datei in Presto.
presto-connector-redis	Ändert die Werte in der <code>redis.properties</code> -Datei in Presto.
presto-connector-tpch	Ändert die Werte in der <code>tpch.properties</code> -Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.

Klassifizierungen	Beschreibung
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.1.0

### 5.1.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#), und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

#### Informationen zur Anwendungsversion

	emr-5.1.0	emr-5.0.3	emr-5.0.2	emr-5.0.1
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.1.3	-	-	-

	emr-5.1.0	emr-5.0.3	emr-5.0.2	emr-5.0.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.3	1.2.2	1.2.2	1.2.2
HCatalog	2.1.0	2.1.0	2.1.0	2.1.0
Hadoop	2.7.3	2.7.3	2.7.2	2.7.2
Hive	2.1.0	2.1.0	2.1.0	2.1.0
Hudi	-	-	-	-
Hue	3.10.0	3.10.0	3.10.0	3.10.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.16.0	0,16,0	0,16,0	0,16,0
Presto	0.152.3	0.152.3	0.150	0.150
Spark	2.0.1	2.0.1	2.0.0	2.0.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-

	emr-5.1.0	emr-5.0.3	emr-5.0.2	emr-5.0.1
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.6.2	0.6.1	0.6.1	0.6.1
ZooKeeper	3.4.8	3.4.8	3.4.8	3.4.8

## 5.1.0 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon EMR Version 5.1.0. Änderungen beziehen sich auf Amazon-EMR-Version 5.0.3.

Veröffentlichungsdatum: 03. November 2016

### Änderungen und Verbesserungen

- Flink 1.1.3 wird jetzt unterstützt.
- Presto wurde als Option im Notebookabschnitt von Hue hinzugefügt.

### Upgrades

- Upgrade auf Hbase 1.2.3
- Upgrade auf Zeppelin 0.6.2

### Bekannte Probleme aus den früheren Versionen, die behoben wurden

- Ein Problem mit Tez-Abfragen in Amazon S3, bei dem ORC-Dateien nicht so gute Leistung zeigten, wie in früheren Amazon-EMR-4.x-Versionen, wurde behoben.

## 5.1.0 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der

Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-amzn-*EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>emr-ddb</code>	4.1.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.1.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.2.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-s3-dist-cp</code>	2.4.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
<code>emrfs</code>	2.11.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>flink-client</code>	1.1.3	Apache Flink-Clientskripts und -Anwendungen für die Befehlszeile.
<code>ganglia-monitor</code>	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem

Komponente	Version	Beschreibung
		zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-0	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-0	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-0	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.7.3-amzn-0	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httpfs-server	2.7.3-amzn-0	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.7.3-amzn-0	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider

Komponente	Version	Beschreibung
hadoop-mapred	2.7.3-amzn-0	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-0	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.3-amzn-0	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-0	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.2.3	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.2.3	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.2.3	HBase-Befehlszeilen-Client.
hbase-rest-server	1.2.3	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.2.3	Service, der einen Thrift-Endpunkt für HBase bereitstellt.



Komponente	Version	Beschreibung
hcatalog-client	2.1.0-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.1.0-amzn-0	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.1.0-amzn-0	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.1.0-amzn-0	Hive-Befehlszeilen-Client.
hive-metastore-server	2.1.0-amzn-0	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server	2.1.0-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.10.0-amzn-0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.52	MySQL-Datenbankserver.
oozie-client	4.2.0	Oozie-Befehlszeilen-Client.

Komponente	Version	Beschreibung
oozie-server	4.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.7.0-HBase-1.2	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.7.0-HBase-1.2	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.152.3	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.152.3	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.16.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	2.0.1	Spark-Befehlszeilen-Clients.
spark-history-server	2.0.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.0.1	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.0.1	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.

Komponente	Version	Beschreibung
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.23	Apache HTTP-Server.
zeppelin-server	0.6.2	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.8	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.8	ZooKeeper Befehlszeilenclient.

## 5.1.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

### emr-5.1.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.

Klassifizierungen	Beschreibung
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
flink-conf	Ändert die flink-conf.yaml-Einstellungen.
flink-log4j	Ändert die log4j.properties-Einstellungen für Flink.
flink-log4j-yarn-session	Ändern Sie die Einstellungen von Flink log4 j-yarn-session .properties.
flink-log4j-cli	Ändert die log4j-cli.properties-Einstellungen für Flink.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.

Klassifizierungen	Beschreibung
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.

Klassifizierungen	Beschreibung
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie die Werte in der Anwendungsumgebung MapReduce .
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.

Klassifizierungen	Beschreibung
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.



Klassifizierungen	Beschreibung
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.0.3

### 5.0.3 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

### Informationen zur Anwendungsversion

	emr-5.0.3	emr-5.0.2	emr-5.0.1	emr-5.0.0
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	2.11,8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.2	1.2.2
HCatalog	2.1.0	2.1.0	2.1.0	2.1.0
Hadoop	2.7.3	2.7.2	2.7.2	2.7.2
Hive	2.1.0	2.1.0	2.1.0	2.1.0
Hudi	-	-	-	-
Hue	3.10.0	3.10.0	3.10.0	3.10.0
Iceberg	-	-	-	-

	emr-5.0.3	emr-5.0.2	emr-5.0.1	emr-5.0.0
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.16.0	0,16,0	0,16,0	0,16,0
Presto	0.152.3	0.150	0.150	0.150
Spark	2.0.1	2.0.0	2.0.0	2.0.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.6.1	0.6.1	0.6.1	0.6.1
ZooKeeper	3.4.8	3.4.8	3.4.8	3.4.8

### 5.0.3 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon EMR Version 5.0.3. Änderungen beziehen sich auf Amazon EMR Version 5.0.0.

Veröffentlichungsdatum: 24. Oktober 2016

## Upgrades

- Upgrade auf Hadoop 2.7.3
- Es wurde ein Upgrade auf Presto 0.152.3 durchgeführt, das Support für die Presto-Webschnittstelle umfasst. Sie können auf die Presto-Webschnittstelle des Presto-Koordinators über Port 8889 zugreifen. Weitere Informationen über die Presto-Webschnittstelle finden Sie unter [Web Interface](#) in der Presto-Dokumentation.
- Upgrade auf Spark 2.0.1
- Amazon-EMR-Versionen basieren jetzt auf Amazon Linux 2016.09. Weitere Informationen finden Sie unter <https://aws.amazon.com/amazon-linux-ami/2016.09-release-notes/>.

### 5.0.3 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>emr-ddb</code>	4.1.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.1.0	Praktische Bibliotheken für das Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-kinesis	3.2.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.4.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.10.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-0	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-0	HDFS-Service auf Knotenebene zum Speichern von Blöcken.

Komponente	Version	Beschreibung
hadoop-hdfs-library	2.7.3-amzn-0	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.7.3-amzn-0	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httfs-server	2.7.3-amzn-0	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.7.3-amzn-0	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.7.3-amzn-0	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-0	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.3-amzn-0	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-0	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.

Komponente	Version	Beschreibung
hbase-hmaster	1.2.2	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.2.2	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.2.2	HBase-Befehlszeilen-Client.
hbase-rest-server	1.2.2	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.2.2	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.1.0-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.1.0-amzn-0	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.1.0-amzn-0	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.1.0-amzn-0	Hive-Befehlszeilen-Client.

Komponente	Version	Beschreibung
hive-metastore-server	2.1.0-amzn-0	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server	2.1.0-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.10.0-amzn-0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.52	MySQL-Datenbankserver.
oozie-client	4.2.0	Oozie-Befehlszeilen-Client.
oozie-server	4.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.7.0-HBase-1.2	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.7.0-HBase-1.2	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.



Komponente	Version	Beschreibung
presto-coordinator	0.152.3	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.152.3	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.16.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	2.0.1	Spark-Befehlszeilen-Clients.
spark-history-server	2.0.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.0.1	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.0.1	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.23	Apache HTTP-Server.
zeppelin-server	0.6.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.

Komponente	Version	Beschreibung
zookeeper-server	3.4.8	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.8	ZooKeeper Befehlszeilenclient.

### 5.0.3 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

#### emr-5.0.3-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.

Klassifizierungen	Beschreibung
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei proto-hive-site.xml von HCatalog.
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.

Klassifizierungen	Beschreibung
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.

Klassifizierungen	Beschreibung
mapred-env	Ändern Sie die Werte in der Anwendungsumgebung MapReduce .
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.

Klassifizierungen	Beschreibung
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.

Klassifizierungen	Beschreibung
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.0.2

### 5.0.2 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

### Informationen zur Anwendungsversion

	emr-5.0.3	emr-5.0.2	emr-5.0.1	emr-5.0.0
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	2.11,8	2.11.8	2.11.8	2.11.8



	emr-5.0.3	emr-5.0.2	emr-5.0.1	emr-5.0.0
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.2	1.2.2
HCatalog	2.1.0	2.1.0	2.1.0	2.1.0
Hadoop	2.7.3	2.7.2	2.7.2	2.7.2
Hive	2.1.0	2.1.0	2.1.0	2.1.0
Hudi	-	-	-	-
Hue	3.10.0	3.10.0	3.10.0	3.10.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.16.0	0,16,0	0,16,0	0,16,0

	emr-5.0.3	emr-5.0.2	emr-5.0.1	emr-5.0.0
Presto	0.152.3	0.150	0.150	0.150
Spark	2.0.1	2.0.0	2.0.0	2.0.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.6.1	0.6.1	0.6.1	0.6.1
ZooKeeper	3.4.8	3.4.8	3.4.8	3.4.8

## 5.0.2 Versionshinweise

### 5.0.2 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
emr-ddb	4.0.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.1.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.2.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.4.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.9.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.

Komponente	Version	Beschreibung
hadoop-client	2.7.2-amzn-3	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.2-amzn-3	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.2-amzn-3	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.7.2-amzn-3	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httpfs-server	2.7.2-amzn-3	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.7.2-amzn-3	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.7.2-amzn-3	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.2-amzn-3	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.2-amzn-3	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.

Komponente	Version	Beschreibung
hadoop-yarn-timeline-server	2.7.2-amzn-3	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.2.2	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.2.2	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.2.2	HBase-Befehlszeilen-Client.
hbase-rest-server	1.2.2	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.2.2	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.1.0-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.1.0-amzn-0	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.1.0-amzn-0	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.

Komponente	Version	Beschreibung
hive-client	2.1.0-amzn-0	Hive-Befehlszeilen-Client.
hive-metastore-server	2.1.0-amzn-0	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server	2.1.0-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.10.0-amzn-0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.46	MySQL-Datenbankserver.
oozie-client	4.2.0	Oozie-Befehlszeilen-Client.
oozie-server	4.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.7.0-HBase-1.2	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.7.0-HBase-1.2	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.

Komponente	Version	Beschreibung
presto-coordinator	0.150	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.150	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.16.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	2.0.0	Spark-Befehlszeilen-Clients.
spark-history-server	2.0.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.0.0	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.0.0	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.23	Apache HTTP-Server.
zeppelin-server	0.6.1-SNAPSHOT	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.

Komponente	Version	Beschreibung
zookeeper-server	3.4.8	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.8	ZooKeeper Befehlszeilenclient.

## 5.0.2 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

### emr-5.0.2-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.



Klassifizierungen	Beschreibung
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei proto-hive-site.xml von HCatalog.
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.

Klassifizierungen	Beschreibung
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.

Klassifizierungen	Beschreibung
mapred-env	Ändern Sie die Werte in der Anwendungsumgebung MapReduce .
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.

Klassifizierungen	Beschreibung
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraooop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.

Klassifizierungen	Beschreibung
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.0.1

### 5.0.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

#### Informationen zur Anwendungsversion

	emr-5.0.3	emr-5.0.2	emr-5.0.1	emr-5.0.0
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	2.11,8	2.11.8	2.11.8	2.11.8

	emr-5.0.3	emr-5.0.2	emr-5.0.1	emr-5.0.0
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.2	1.2.2
HCatalog	2.1.0	2.1.0	2.1.0	2.1.0
Hadoop	2.7.3	2.7.2	2.7.2	2.7.2
Hive	2.1.0	2.1.0	2.1.0	2.1.0
Hudi	-	-	-	-
Hue	3.10.0	3.10.0	3.10.0	3.10.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.16.0	0,16,0	0,16,0	0,16,0

	emr-5.0.3	emr-5.0.2	emr-5.0.1	emr-5.0.0
Presto	0.152.3	0.150	0.150	0.150
Spark	2.0.1	2.0.0	2.0.0	2.0.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.6.1	0.6.1	0.6.1	0.6.1
ZooKeeper	3.4.8	3.4.8	3.4.8	3.4.8

## 5.0.1 Versionshinweise

### 5.0.1 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.



Komponente	Version	Beschreibung
emr-ddb	4.0.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.1.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.2.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.4.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.9.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.

Komponente	Version	Beschreibung
hadoop-client	2.7.2-amzn-3	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.2-amzn-3	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.2-amzn-3	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.7.2-amzn-3	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httpfs-server	2.7.2-amzn-3	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.7.2-amzn-3	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.7.2-amzn-3	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.2-amzn-3	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.2-amzn-3	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.

Komponente	Version	Beschreibung
hadoop-yarn-timeline-server	2.7.2-amzn-3	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.2.2	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.2.2	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.2.2	HBase-Befehlszeilen-Client.
hbase-rest-server	1.2.2	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.2.2	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	2.1.0-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.1.0-amzn-0	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.1.0-amzn-0	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.

Komponente	Version	Beschreibung
hive-client	2.1.0-amzn-0	Hive-Befehlszeilen-Client.
hive-metastore-server	2.1.0-amzn-0	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server	2.1.0-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.10.0-amzn-0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.46	MySQL-Datenbankserver.
oozie-client	4.2.0	Oozie-Befehlszeilen-Client.
oozie-server	4.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.7.0-HBase-1.2	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.7.0-HBase-1.2	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.

Komponente	Version	Beschreibung
presto-coordinator	0.150	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.150	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.16.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	2.0.0	Spark-Befehlszeilen-Clients.
spark-history-server	2.0.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.0.0	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.0.0	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.23	Apache HTTP-Server.
zeppelin-server	0.6.1-SNAPSHOT	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.

Komponente	Version	Beschreibung
zookeeper-server	3.4.8	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.8	ZooKeeper Befehlszeilenclient.

## 5.0.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

### emr-5.0.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.

Klassifizierungen	Beschreibung
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei proto-hive-site.xml von HCatalog.
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.

Klassifizierungen	Beschreibung
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.



Klassifizierungen	Beschreibung
mapred-env	Ändern Sie die Werte in der Anwendungsumgebung MapReduce .
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.

Klassifizierungen	Beschreibung
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.

Klassifizierungen	Beschreibung
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 5.0.0

### 5.0.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) und [ZooKeeper](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

### Informationen zur Anwendungsversion

	emr-5.0.3	emr-5.0.2	emr-5.0.1	emr-5.0.0
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	2.11,8	2.11.8	2.11.8	2.11.8

	emr-5.0.3	emr-5.0.2	emr-5.0.1	emr-5.0.0
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.2	1.2.2
HCatalog	2.1.0	2.1.0	2.1.0	2.1.0
Hadoop	2.7.3	2.7.2	2.7.2	2.7.2
Hive	2.1.0	2.1.0	2.1.0	2.1.0
Hudi	-	-	-	-
Hue	3.10.0	3.10.0	3.10.0	3.10.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.16.0	0,16,0	0,16,0	0,16,0

	emr-5.0.3	emr-5.0.2	emr-5.0.1	emr-5.0.0
Presto	0.152.3	0.150	0.150	0.150
Spark	2.0.1	2.0.0	2.0.0	2.0.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.6.1	0.6.1	0.6.1	0.6.1
ZooKeeper	3.4.8	3.4.8	3.4.8	3.4.8

## 5.0.0 Versionshinweise

Veröffentlichungsdatum: 27. Juli 2016

### Upgrades

- Upgrade auf Hive 2.1
- Upgrade auf Presto 0.150
- Upgrade auf Spark 2.0
- Upgrade auf Hue 3.10.0
- Upgrade auf Pig 0.16.0
- Upgrade auf Tez 0.8.4
- Upgrade auf Zeppelin 0.6.1

### Änderungen und Verbesserungen

- Amazon EMR unterstützt die neuesten Open-Source-Versionen von Hive (Version 2.1) und Pig (Version 0.16.0). Wenn Sie Hive oder Pig in der Vergangenheit in Amazon EMR verwendet haben,

kann sich dies auf einige Anwendungsfälle auswirken. Weitere Informationen finden Sie unter [Hive](#) und [Pig](#).

- Die Standard-Ausführungs-Engine für Hive und Pig ist jetzt Tez. Um dies zu ändern, bearbeiten Sie die entsprechenden Werte in den Konfigurationsklassifizierungen `hive-site` bzw. `pig-properties`.
- Eine verbesserte Schritt-Debugging-Funktion wurde hinzugefügt. Mit dieser Funktion können Sie die Ursache von Schrittfehlern sehen, wenn der Service die Ursache ermitteln kann. Weitere Informationen finden Sie unter [Verbessertes Schritt-Debuggen](#) im Verwaltungshandbuch für Amazon EMR.
- Anwendungen, die zuvor mit "-Sandbox" endeten, tragen diesen Suffix nicht mehr. Dies kann Ihre Automatisierung, zum Beispiel, wenn Sie Skripts zum Starten von Clustern mit diesen Anwendungen verwenden, unterbrechen. In der folgenden Tabelle sind die Namen von Anwendungen in Amazon EMR 4.7.2 im Vergleich zu Amazon EMR 5.0.0 aufgeführt.

#### Änderungen der Anwendungsnamen

Amazon EMR 4.7.2	Amazon EMR 5.0.0
Oozie-Sandbox	Oozie
Presto-Sandbox	Presto
Sqoop-Sandbox	Sqoop
Zeppelin-Sandbox	Zeppelin
ZooKeeper-Sandkasten	ZooKeeper

- Spark ist jetzt für Scala 2.11 kompiliert.
- Java 8 ist jetzt die Standard-JVM. Alle Anwendungen werden mit der Java 8-Laufzeit ausgeführt. Es sind keine Änderungen am Byte-Codeziel der Anwendungen vorgenommen worden. Die meisten Anwendungen zielen weiterhin auf Java 7 ab.
- Zeppelin enthält jetzt Authentifizierungsfunktionen. Weitere Informationen finden Sie unter [Zeppelin](#).
- Es werden jetzt Sicherheitskonfigurationen unterstützt, mit deren Hilfe Sie die Verschlüsselungsoptionen einfacher erstellen und anwenden können. Weitere Informationen finden Sie unter [Datenverschlüsselung](#).

## 5.0.0 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>emr-ddb</code>	4.0.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.1.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.2.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-s3-dist-cp</code>	2.4.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
<code>emrfs</code>	2.9.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>ganglia-monitor</code>	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus



Komponente	Version	Beschreibung
		dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.2-amzn-3	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.2-amzn-3	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.2-amzn-3	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.7.2-amzn-3	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httpfs-server	2.7.2-amzn-3	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.7.2-amzn-3	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider

Komponente	Version	Beschreibung
hadoop-mapred	2.7.2-amzn-3	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.2-amzn-3	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.2-amzn-3	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.2-amzn-3	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.2.2	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.2.2	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.2.2	HBase-Befehlszeilen-Client.
hbase-rest-server	1.2.2	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.2.2	Service, der einen Thrift-Endpunkt für HBase bereitstellt.

Komponente	Version	Beschreibung
hcatalog-client	2.1.0-amzn-0	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	2.1.0-amzn-0	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	2.1.0-amzn-0	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	2.1.0-amzn-0	Hive-Befehlszeilen-Client.
hive-metastore-server	2.1.0-amzn-0	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server	2.1.0-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.10.0-amzn-0	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.46	MySQL-Datenbankserver.
oozie-client	4.2.0	Oozie-Befehlszeilen-Client.

Komponente	Version	Beschreibung
oozie-server	4.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.7.0-HBase-1.2	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.7.0-HBase-1.2	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.150	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.150	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.16.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	2.0.0	Spark-Befehlszeilen-Clients.
spark-history-server	2.0.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	2.0.0	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	2.0.0	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.

Komponente	Version	Beschreibung
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.23	Apache HTTP-Server.
zeppelin-server	0.6.1-SNAPSHOT	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.8	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.8	ZooKeeper Befehlszeilenclient.

## 5.0.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

### emr-5.0.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.

Klassifizierungen	Beschreibung
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.

Klassifizierungen	Beschreibung
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei proto-hive-site.xml von HCatalog.
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j2	Ändert die Werte in der log4j2.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-beeline-log4j2	Ändert die Werte in der beeline-log4j2.properties-Datei in Hive.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von hive-exec-log Hive.
hive-llap-daemon-log4j2	Ändern Sie die Werte in der Datei 4j2.properties von Hive. llap-daemon-log
hive-log4j2	Ändert die Werte in der hive-log4j2.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.

Klassifizierungen	Beschreibung
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie die Werte in der Anwendungsumgebung MapReduce .
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.



Klassifizierungen	Beschreibung
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-hive-site	Ändert die Werte in der hive-site.xml-Datei in Spark.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.

Klassifizierungen	Beschreibung
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-4.x-Versionen

Dieser Abschnitt enthält Anwendungsversionen, Versionshinweise, Komponentenversionen und Konfigurationsklassifizierungen, die in jeder Amazon-EMR-4.x-Version verfügbar ist.

Wenn Sie einen Cluster starten, können Sie aus verschiedenen Versionen von Amazon EMR wählen. Auf diese Weise können Sie Anwendungsversionen testen und verwenden, die zu den Kompatibilitätsanforderungen passen. Sie geben die Versionsnummer zusammen mit dem Release-Label an. Versionsbezeichnungen haben die Form `emr-x.x.x`. z. B. `emr-7.1.0`.

Neue Amazon-EMR-Versionen werden über einen Zeitraum von mehreren Tagen in verschiedenen Regionen verfügbar gemacht, beginnend mit der ersten Region am ursprünglichen Veröffentlichungsdatum. Die neueste Version ist in diesem Zeitraum möglicherweise nicht in Ihrer Region verfügbar.

Eine umfassende Tabelle der Anwendungsversionen in jeder Amazon-EMR-4.x-Version finden Sie unter [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#).

### Themen

- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)
- [Unterschiede in Amazon-EMR-4.x-Versionen](#)

- [Amazon-EMR-Version 4.9.6](#)
- [Amazon-EMR-Version 4.9.5](#)
- [Amazon-EMR-Version 4.9.4](#)
- [Amazon-EMR-Version 4.9.3](#)
- [Amazon-EMR-Version 4.9.2](#)
- [Amazon-EMR-Version 4.9.1](#)
- [Amazon-EMR-Version 4.8.5](#)
- [Amazon-EMR-Version 4.8.4](#)
- [Amazon-EMR-Version 4.8.3](#)
- [Amazon-EMR-Version 4.8.2](#)
- [Amazon EMR Version 4.8.1](#)
- [Amazon-EMR-Version 4.8.0](#)
- [Amazon-EMR-Version 4.7.4](#)
- [Amazon EMR Version 4.7.3](#)
- [Amazon-EMR-Version 4.7.2](#)
- [Amazon-EMR-Version 4.7.1](#)
- [Amazon-EMR-Version 4.7.0](#)
- [Amazon EMR Version 4.6.1](#)
- [Amazon-EMR-Version 4.6.0](#)
- [Amazon-EMR-Version 4.5.0](#)
- [Amazon-EMR-Version 4.4.0](#)
- [Amazon-EMR-Version 4.3.0](#)
- [Amazon-EMR-Version 4.2.0](#)
- [Amazon-EMR-Version 4.1.0](#)
- [Amazon-EMR-Version 4.0.0](#)

## Anwendungsversionen in Amazon-EMR-4.x-Versionen

Eine umfassende Tabelle mit den in jeder Amazon-EMR-4.x-Version verfügbaren Anwendungsversionen finden Sie unter [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#) in Ihrem Browser.

## Unterschiede in Amazon-EMR-4.x-Versionen

Die Dokumentation für Amazon-EMR-Features in Amazon EMR gibt die Verwaltungshandbuch-Version an, in der ein Feature verfügbar wurde, sowie zutreffende Unterschiede zwischen Amazon-EMR-Features, die auf Version 4.0.0 zurückreichen.

Beginnend mit Amazon-EMR-Version 5.0.0 erhielten einige Anwendungen ein erhebliches Versionsupgrade, bei dem die Installation oder betriebliche Details geändert wurden. Andere wurden von Sandbox-Anwendungen zu nativen Anwendungen hochgestuft. Jedes Thema in diesem Abschnitt bietet bedeutende anwendungsspezifische Unterschiede bei der Verwendung von Amazon-EMR-4.x-Versionen.

### Themen

- [Sandbox-Anwendungen](#)
- [Überlegungen für die Verwendung von Hive in Amazon EMR 4.x](#)
- [Überlegungen für die Verwendung von Pig in Amazon EMR 4.x](#)

### Sandbox-Anwendungen

Bei der Verwendung von Amazon-EMR-4.x-Versionen werden einige Anwendungen als Sandbox-Anwendungen betrachtet. Sandbox-Anwendungen sind frühe Versionen der Anwendung, die wir aufgrund der Nachfrage zum Zeitpunkt der ersten Amazon-EMR-Version zur Verfügung gestellt haben. Sie können die Konsole, AWS CLI oder API installieren, damit Amazon-EMR-Sandbox-Anwendungen wie native Anwendungen installiert. Sandbox-Anwendungen erhalten jedoch eingeschränkten Support und eingeschränkte Dokumentation. Sandbox-Anwendungen wurden in Amazon-EMR-Version 5.0.0 und höher zu nativen, voll unterstützten Anwendungen. Die folgenden Anwendungen sind Sandbox-Anwendungen in Amazon-EMR-4.x-Versionen:

- Oozie
- Presto
- Sqoop
- Zeppelin
- ZooKeeper

Wenn Sie Sandbox-Anwendungen installieren, werden die Anwendungsnamen mit dem Suffix - sandbox angegeben. Beispiel: Zum Installieren der Sandbox-Version von *Presto* verwenden Sie

**Presto-sandbox.** Die Installation kann länger als bei einer voll unterstützten Anwendung dauern. Die für jede Anwendung in diesem Abschnitt aufgelisteten Versionsnummern entsprechen der Community-Version der Anwendung.

## Oozie (Sandbox-Versionen)

Oozie ist als Sandbox-Anwendung ab Amazon-EMR-Version 4.1.0 verfügbar.

Oozie-Beispiele werden von den Sandbox-Versionen nicht standardmäßig installiert. Zum Installieren der Beispiele nutzen Sie SSH für den Master-Knoten und führen Sie `install-oozie-examples` aus.

### Informationen zur Oozie-Sandbox-Version

Amazon-EMR-Version sbezeichnung	Oozie-Sandbox-Version	Mit Oozie-Sandbox installierte Komponenten
emr-4.9.6	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, oozie-client, oozie-server
emr-4.9.5	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager

Amazon-EMR-Version sbezeichnung	Oozie-Sandbox-Version	Mit Oozie-Sandbox installierte Komponenten
		cemanager, oozie-client, oozie-server
emr-4.9.4	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcecemanager, oozie-client, oozie-server
emr-4.9.3	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcecemanager, oozie-client, oozie-server

Amazon-EMR-Version sbezeichnung	Oozie-Sandbox-Version	Mit Oozie-Sandbox installierte Komponenten
emr-4.9.2	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, oozie-client, oozie-server
emr-4.9.1	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, oozie-client, oozie-server



Amazon-EMR-Version sbezeichnung	Oozie-Sandbox-Version	Mit Oozie-Sandbox installierte Komponenten
emr-4.8.5	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, oozie-client, oozie-server
emr-4.8.4	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, oozie-client, oozie-server

Amazon-EMR-Version sbezeichnung	Oozie-Sandbox-Version	Mit Oozie-Sandbox installierte Komponenten
emr-4.8.3	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, oozie-client, oozie-server
emr-4.8.2	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, oozie-client, oozie-server

Amazon-EMR-Version sbezeichnung	Oozie-Sandbox-Version	Mit Oozie-Sandbox installierte Komponenten
emr-4.8.1	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, oozie-client, oozie-server
emr-4.8.0	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, oozie-client, oozie-server

Amazon-EMR-Version sbezeichnung	Oozie-Sandbox-Version	Mit Oozie-Sandbox installierte Komponenten
emr-4.7.4	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, oozie-client, oozie-server
emr-4.7.3	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, oozie-client, oozie-server

Amazon-EMR-Version sbezeichnung	Oozie-Sandbox-Version	Mit Oozie-Sandbox installierte Komponenten
emr-4.7.2	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, oozie-client, oozie-server
emr-4.7.1	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, oozie-client, oozie-server

Amazon-EMR-Version sbezeichnung	Oozie-Sandbox-Version	Mit Oozie-Sandbox installierte Komponenten
emr-4.7.0	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server
emr-4.6.1	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server
emr-4.6.0	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server

Amazon-EMR-Version sbezeichnung	Oozie-Sandbox-Version	Mit Oozie-Sandbox installierte Komponenten
emr-4.5.0	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httplibfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server
emr-4.4.0	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httplibfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server
emr-4.3.0	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httplibfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server

Amazon-EMR-Version sbezeichnung	Oozie-Sandbox-Version	Mit Oozie-Sandbox installierte Komponenten
emr-4.2.0	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httplib-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server
emr-4.1.0	4.0.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httplib-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server

## Presto (Sandbox-Versionen)

Presto ist als Sandbox-Anwendung ab Amazon-EMR-Version 4.1.0 verfügbar.

### Informationen zur Presto-Sandbox-Version

Amazon-EMR-Version sbezeichnung	Presto-Sandbox-Version	Mit Presto-Sandbox installierte Komponenten
emr-4.9.6	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-



Amazon-EMR-Version sbezeichnung	Presto-Sandbox-Version	Mit Presto-Sandbox installierte Komponenten
		kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-4.9.5	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-4.9.4	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon-EMR-Version sbezeichnung	Presto-Sandbox-Version	Mit Presto-Sandbox installierte Komponenten
emr-4.9.3	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-4.9.2	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-4.9.1	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon-EMR-Version sbezeichnung	Presto-Sandbox-Version	Mit Presto-Sandbox installierte Komponenten
emr-4.8.5	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-4.8.4	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-4.8.3	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon-EMR-Version sbezeichnung	Presto-Sandbox-Version	Mit Presto-Sandbox installierte Komponenten
emr-4.8.2	0.152.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-4.8.1	0.151	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-4.8.0	0.151	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon-EMR-Version sbezeichnung	Presto-Sandbox-Version	Mit Presto-Sandbox installierte Komponenten
emr-4.7.4	0.148	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-4.7.3	0.148	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-4.7.2	0.148	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon-EMR-Version sbezeichnung	Presto-Sandbox-Version	Mit Presto-Sandbox installierte Komponenten
emr-4.7.1	0.147	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-4.7.0	0.147	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-4.6.1	0.143	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, mysql-server, presto-coordinator, presto-worker

Amazon-EMR-Version sbezeichnung	Presto-Sandbox-Version	Mit Presto-Sandbox installierte Komponenten
emr-4.6.0	0.143	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hive-client, hive-metastore-server, mysql-server, presto-coordinator, presto-worker
emr-4.5.0	0.140	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hive-client, hive-metastore-server, mysql-server, presto-coordinator, presto-worker
emr-4.4.0	0.136	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hive-client, hive-metastore-server, mysql-server, presto-coordinator, presto-worker

Amazon-EMR-Version sbezeichnung	Presto-Sandbox-Version	Mit Presto-Sandbox installierte Komponenten
emr-4.3.0	0.130	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, mysql-server, presto-coordinator, presto-worker
emr-4.2.0	0.125	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, mysql-server, presto-coordinator, presto-worker
emr-4.1.0	0.119	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, mysql-server, presto-coordinator, presto-worker

### Sqoop (Sandbox-Versionen)

Sqoop ist als Sandbox-Anwendung ab Amazon-EMR-Version 4.4.0 verfügbar.



## Informationen zur Sqoop-Sandbox-Version

Amazon-EMR-Version sbezeichnung	Sqoop-Sandbox-Version	Mit Sqoop-Sandbox installierte Komponenten
emr-4.9.6	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-4.9.5	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-4.9.4	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client

Amazon-EMR-Version sbezeichnung	Sqoop-Sandbox-Version	Mit Sqoop-Sandbox installierte Komponenten
emr-4.9.3	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-4.9.2	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-4.9.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client

Amazon-EMR-Version sbezeichnung	Sqoop-Sandbox-Version	Mit Sqoop-Sandbox installierte Komponenten
emr-4.8.5	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-4.8.4	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-4.8.3	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client

Amazon-EMR-Version sbezeichnung	Sqoop-Sandbox-Version	Mit Sqoop-Sandbox installierte Komponenten
emr-4.8.2	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-4.8.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-4.8.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client

Amazon-EMR-Version sbezeichnung	Sqoop-Sandbox-Version	Mit Sqoop-Sandbox installierte Komponenten
emr-4.7.4	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-4.7.3	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-4.7.2	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client

Amazon-EMR-Version sbezeichnung	Sqoop-Sandbox-Version	Mit Sqoop-Sandbox installierte Komponenten
emr-4.7.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-4.7.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-4.6.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client

Amazon-EMR-Version sbezeichnung	Sqoop-Sandbox-Version	Mit Sqoop-Sandbox installierte Komponenten
emr-4.6.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httplibfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-4.5.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httplibfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, sqoop-client
emr-4.4.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httplibfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, sqoop-client

### Zeppelin (Sandbox-Versionen)

Zeppelin ist als Sandbox-Anwendung ab Amazon-EMR-Version 4.1.0 verfügbar.

## Informationen zur Zeppelin-Sandbox-Version

Amazon-EMR-Version sbezeichnung	Zeppelin-Sandbox-Versionen	Mit Zeppelin-Sandbox installierte Komponenten
emr-4.9.6	0.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-4.9.5	0.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-4.9.4	0.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-



Amazon-EMR-Version sbezeichnung	Zeppelin-Sandbox-Versionen	Mit Zeppelin-Sandbox installierte Komponenten
		on-yarn, spark-yarn-slave, zeppelin-server
emr-4.9.3	0.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-4.9.2	0.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon-EMR-Version sbezeichnung	Zeppelin-Sandbox-Versionen	Mit Zeppelin-Sandbox installie rte Komponenten
emr-4.9.1	0.6.1	emrfs, emr-goodies, hadoop- client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, spark-client, spark-history-server, spark- on-yarn, spark-yarn-slave, zeppelin-server
emr-4.8.5	0.6.1	emrfs, emr-goodies, hadoop- client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, spark-client, spark-history-server, spark- on-yarn, spark-yarn-slave, zeppelin-server

Amazon-EMR-Version sbezeichnung	Zeppelin-Sandbox-Versionen	Mit Zeppelin-Sandbox installie rte Komponenten
emr-4.8.4	0.6.1	emrfs, emr-goodies, hadoop- client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, spark-client, spark-history-server, spark- on-yarn, spark-yarn-slave, zeppelin-server
emr-4.8.3	0.6.1	emrfs, emr-goodies, hadoop- client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, spark-client, spark-history-server, spark- on-yarn, spark-yarn-slave, zeppelin-server

Amazon-EMR-Version sbezeichnung	Zeppelin-Sandbox-Versionen	Mit Zeppelin-Sandbox installie rte Komponenten
emr-4.8.2	0.6.1	emrfs, emr-goodies, hadoop- client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, spark-client, spark-history-server, spark- on-yarn, spark-yarn-slave, zeppelin-server
emr-4.8.1	0.6.1	emrfs, emr-goodies, hadoop- client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, spark-client, spark-history-server, spark- on-yarn, spark-yarn-slave, zeppelin-server

Amazon-EMR-Version sbezeichnung	Zeppelin-Sandbox-Versionen	Mit Zeppelin-Sandbox installie rte Komponenten
emr-4.8.0	0.6.1	emrfs, emr-goodies, hadoop- client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, spark-client, spark-history-server, spark- on-yarn, spark-yarn-slave, zeppelin-server
emr-4.7.4	0.5.6	emrfs, emr-goodies, hadoop- client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, spark-client, spark-history-server, spark- on-yarn, spark-yarn-slave, zeppelin-server

Amazon-EMR-Version sbezeichnung	Zeppelin-Sandbox-Versionen	Mit Zeppelin-Sandbox installie rte Komponenten
emr-4.7.3	0.5.6	emrfs, emr-goodies, hadoop- client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, spark-client, spark-history-server, spark- on-yarn, spark-yarn-slave, zeppelin-server
emr-4.7.2	0.5.6	emrfs, emr-goodies, hadoop- client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, spark-client, spark-history-server, spark- on-yarn, spark-yarn-slave, zeppelin-server

Amazon-EMR-Version sbezeichnung	Zeppelin-Sandbox-Versionen	Mit Zeppelin-Sandbox installierte Komponenten
emr-4.7.1	0.5.6	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-4.7.0	0.5.6	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-4.6.1	0.5.6	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon-EMR-Version sbezeichnung	Zeppelin-Sandbox-Versionen	Mit Zeppelin-Sandbox installierte Komponenten
emr-4.6.0	0.5.6	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ht tpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resource manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-4.5.0	0.5.6	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ht tpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resource manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-4.4.0	0.5.6	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ht tpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resource manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server



Amazon-EMR-Version sbezeichnung	Zeppelin-Sandbox-Versionen	Mit Zeppelin-Sandbox installierte Komponenten
emr-4.3.0	0.5.5	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ht tpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resource manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-4.2.0	0.5.5	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ht tpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resource manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-4.1.0	0.6.0-SNAPSHOT	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ht tpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resource manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

## Zookeeper (Sandbox-Versionen)

Zookeeper ist als Sandbox-Anwendung ab Amazon-EMR-Version 4.6.0 verfügbar.

### Informationen zur Version von Zookeeper-Sandbox

Amazon-EMR-Version sbezeichnung	Zookeeper-Sandbox-Version	Mit ZooKeeper-Sandbox installierte Komponenten
emr-4.9.6	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server
emr-4.9.5	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server
emr-4.9.4	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Zookeeper-Sandbox-Version	Mit ZooKeeper-Sandbox installierte Komponenten
emr-4.9.3	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server
emr-4.9.2	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server
emr-4.9.1	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Zookeeper-Sandbox-Version	Mit ZooKeeper-Sandbox installierte Komponenten
emr-4.8.5	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server
emr-4.8.4	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server
emr-4.8.3	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Zookeeper-Sandbox-Version	Mit ZooKeeper-Sandbox installierte Komponenten
emr-4.8.2	3.4.8	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server
emr-4.8.1	3.4.8	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server
emr-4.8.0	3.4.8	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Zookeeper-Sandbox-Version	Mit ZooKeeper-Sandbox installierte Komponenten
emr-4.7.4	3.4.8	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server
emr-4.7.3	3.4.8	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server
emr-4.7.2	3.4.8	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Zookeeper-Sandbox-Version	Mit ZooKeeper-Sandbox installierte Komponenten
emr-4.7.1	3.4.8	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server
emr-4.7.0	3.4.8	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server
emr-4.6.1	3.4.8	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Zookeeper-Sandbox-Version	Mit ZooKeeper-Sandbox installierte Komponenten
emr-4.6.0	3.4.8	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ht tpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server

## Überlegungen für die Verwendung von Hive in Amazon EMR 4.x

In diesem Abschnitt werden die Unterschiede behandelt, die Sie berücksichtigen sollten, wenn Sie Hive-Version 1.0.0 auf Amazon-EMR-4.x-Versionen im Vergleich zu Hive 2.x auf Amazon-EMR-5.x-Versionen verwenden.

### ACID-Transaktionen nicht unterstützt

Hive auf Amazon-EMR-4.x-Versionen unterstützt keine ACID-Transaktionen mit Hive-Daten, die in Amazon S3 gespeichert werden, wenn 4.x-Versionen verwendet werden. Wenn Sie versuchen, eine transaktionale Tabelle in Amazon S3 zu erstellen, tritt eine Ausnahme auf.

### Lesen und Schreiben von Tabellen in Amazon S3

Hive in Amazon-EMR-4.x-Versionen können direkt ohne die Verwendung temporärer Dateien in Amazon S3 schreiben. Dadurch wird die Leistung verbessert, eine Folge davon ist allerdings, dass Sie nicht in derselben Tabelle in Amazon S3 innerhalb derselben Hive-Anweisung lesen und schreiben können. Sie können dieses Problem umgehen, indem Sie eine temporäre Tabelle in HDFS erstellen und verwenden.

Das folgende Beispiel zeigt, wie Sie mit mehreren Hive-Anweisungen eine Tabelle in Amazon S3 aktualisieren. Die Anweisungen erstellen eine temporäre Tabelle in HDFS mit dem Namen tmp auf Grundlage einer Tabelle in Amazon S3 mit dem Namen my\_s3\_table. Die Tabelle in Amazon S3 wird dann mit den Inhalten der temporären Tabelle aktualisiert.



```
CREATE TEMPORARY TABLE tmp LIKE my_s3_table;  
INSERT OVERWRITE TABLE tmp SELECT ....;  
INSERT OVERWRITE TABLE my_s3_table SELECT * FROM tmp;
```

## Log4j im Vergleich zu Log4j 2

Hive auf Amazon-EMR-4.x-Versionen verwendet Log4j. Ab Version 5.0.0, Log4j 2 ist die Standardeinstellung. Diese Versionen erfordern möglicherweise unterschiedliche Protokollierungskonfigurationen. Weitere Informationen finden Sie unter [Apache Log4j 2](#).

## MapReduce ist die Standard-Ausführungs-Engine

Hive auf Amazon-EMR-4.x-Versionen verwendet MapReduce als Standard-Ausführungs-Engine. Ab Amazon-EMR-Version 5.0.0 ist Tez die Standardversion, die für die meisten Workflows eine verbesserte Leistung bietet.

## Hive-Autorisierung

Hive auf Amazon EMR 4.x-Versionen unterstützt die [Hive-Autorisierung](#) für HDFS, aber nicht für EMRFS und Amazon S3. Amazon-EMR-Cluster werden standardmäßig mit deaktivierter Autorisierung ausgeführt.

## Dateizusammenführung mit Hive in Amazon S3

Hive auf Amazon-EMR-4.x-Versionen führt kleine Dateien am Ende eines Map-Auftrags zusammen, wenn `hive.merge.mapfiles true` ist. Eine Zusammenführung wird nur ausgelöst, wenn die durchschnittliche Leistungsgröße des Auftrags kleiner ist als die `hive.merge.smallfiles.avgsize`-Einstellung. Amazon EMR; Hive verhält sich genau gleich, wenn sich der Endausgabepfad in HDFS befindet. Wenn sich der Ausgabepfad in Amazon S3 befindet, wird der Parameter `hive.merge.smallfiles.avgsize` jedoch ignoriert. In diesem Fall wird die Zusammenführungsaufgabe immer dann ausgelöst, wenn `hive.merge.mapfiles` auf `true` gesetzt ist.

## Überlegungen für die Verwendung von Pig in Amazon EMR 4.x

Pig-Version 0.14.0 ist auf Clustern installiert, die mit Amazon-EMR-4.x-Versionen erstellt wurden. Für Pig wurde in Amazon EMR 5.0.0 ein Upgrade auf Version 0.16.0 durchgeführt. Die wesentlichen Unterschiede werden unten erläutert.

## Unterschiedliche Standard-Ausführungs-Engine

Pig-Version 0.14.0 auf Amazon-EMR-4.x-Versionen verwendet MapReduce als Standard-Ausführungs-Engine. Pig 0.16.0 und höher verwenden Apache Tez. Sie können explizit `execType=mapreduce` in der Konfigurationsklassifizierung `pig-properties` festlegen, um MapReduce zu verwenden.

## Gelöschte benutzerdefinierte Pig-Funktionen (UDFs)

Benutzerdefinierte UDFs, die in Pig auf Amazon-EMR-4.x-Versionen waren, waren ab Pig 0.16.0 nicht mehr enthalten. Die meisten UDFs weisen ähnliche Funktionen auf, die Sie stattdessen verwenden können. Die folgende Tabelle listet gelöschte UDFs und ähnliche Funktionen auf. Weitere Informationen finden Sie unter [Integrierte Funktionen](#) auf der Apache-Pig-Website.

Gelöschte UDF	Gleichwertige Funktion
FORMAT_DT(dtformat, date)	GetHour(date), GetMinute(date), GetMonth(date), GetSecond(date), GetWeek(date), GetYear(date), GetDay(date)
EXTRACT(string, pattern)	REGEX_EXTRACT_ALL(string, pattern)
REPLACE(string, pattern, replacement)	REPLACE(string, pattern, replacement)
DATE_TIME()	ToDate()
DURATION(dt, dt2)	WeeksBetween(dt, dt2), YearsBetween(dt, dt2), SecondsBetween(dt, dt2), MonthsBetween(dt, dt2), MinutesBetween(dt, dt2), HoursBetween(dt, dt2)
EXTRACT_DT(format, date)	GetHour(date), GetMinute(date), GetMonth(date), GetSecond(date), GetWeek(date), GetYear(date), GetDay(date)
OFFSET_DT(date, duration)	AddDuration(date, duration), SubtractDuration(date, duration)
PERIOD(dt, dt2)	WeeksBetween(dt, dt2), YearsBetween(dt, dt2), SecondsBetween(dt, dt2), MonthsBetween(dt,

Gelöschte UDF	Gleichwertige Funktion
	dt2), MinutesBetween(dt, dt2), HoursBetween(dt, dt2)
CAPITALIZE(string)	UCFIRST(string)
CONCAT_WITH()	CONCAT()
INDEX_OF()	INDEXOF()
LAST_INDEX_OF()	LAST_INDEXOF()
SPLIT_ON_REGEX()	STRSPLT()
UNCAPITALIZE()	LCFIRST()

Die folgenden UDFs wurden ohne gleichwertige Funktionen entfernt: FORMAT(), LOCAL\_DATE(), LOCAL\_TIME(), CENTER(), LEFT\_PAD(), REPEAT(), REPLACE\_ONCE(), RIGHT\_PAD(), STRIP(), STRIIP\_END(), STRIP\_START(), SWAP\_CASE().

#### Nicht mehr enthaltene Grunt-Befehle

Einige Grunt-Befehle werden ab Pig 0.16.0 nicht mehr angeboten. In der folgenden Tabelle sind die Grunt-Befehle in Pig 0.14.0 und ihre entsprechenden Befehle (sofern vorhanden) in der aktuellen Version aufgeführt.

#### Pig 0.14.0 und gleichwertige aktuelle Grunt-Befehle

Grunt-Befehl in Pig 0.14.0	Pig Grunt-Befehl in 0.16.0 und höher
cat <non-hdfs-path>)	fs -cat <non-hdfs-path>;
cd <non-hdfs-path>;	Keine Entsprechung
ls <non-hdfs-path>;	fs -ls <non-hdfs-path>;
move <non-hdfs-path> <non-hdfs-path>;	fs -mv <non-hdfs-path> <non-hdfs-path>;
copy <non-hdfs-path> <non-hdfs-path>;	fs -cp <non-hdfs-path> <non-hdfs-path>;

Grunt-Befehl in Pig 0.14.0	Pig Grunt-Befehl in 0.16.0 und höher
<code>copyToLocal &lt;non-hdfs-path&gt; &lt;local-path&gt;;</code>	<code>fs -copyToLocal &lt;non-hdfs-path&gt; &lt;local-path&gt;;</code>
<code>copyFromLocal &lt;local-path&gt; &lt;non-hdfs-path&gt;;</code>	<code>fs -copyFromLocal &lt;local-path&gt; &lt;non-hdfs-path&gt;;</code>
<code>mkdir &lt;non-hdfs-path&gt;;</code>	<code>fs -mkdir &lt;non-hdfs-path&gt;;</code>
<code>rm &lt;non-hdfs-path&gt;;</code>	<code>fs -rm -r -skipTrash &lt;non-hdfs-path&gt;;</code>
<code>rmf &lt;non-hdfs-path&gt;;</code>	<code>fs -rm -r -skipTrash &lt;non-hdfs-path&gt;;</code>

## Entfernte Funktionen für Nicht-HDFS-Stammverzeichnisse

Pig 0.14.0 in Amazon-EMR-4.x-Versionen verfügt über zwei Mechanismen, um anderen Benutzern als dem `hadoop`-Benutzer, die keine Root-Verzeichnisse besitzen, das Ausführen von Pig-Skripten zu ermöglichen. Der erste Mechanismus ist ein automatisches Fallback, mit dem das anfängliche Arbeitsverzeichnis auf das Stammverzeichnis festgelegt wird, wenn kein Stammverzeichnis vorhanden ist. Der zweite ist eine `pig.initial.fs.name`-Eigenschaft, über die Sie das anfängliche Arbeitsverzeichnis ändern können.

Diese Mechanismen sind beginnend mit Amazon-EMR-Version 5.0.0 nicht verfügbar, und Benutzer müssen ein Stammverzeichnis in HDFS haben. Dies gilt nicht für `hadoop`-Benutzer, da hier ein Stammverzeichnis beim Starten bereitgestellt wird. Bei Skriptläufen, bei denen Hadoop-JAR-Schritte verwendet werden, wird standardmäßig der Hadoop-Benutzer vorgeschlagen, es sei denn, es ist explizit ein anderer Benutzer über `command-runner.jar` angegeben.

## Amazon-EMR-Version 4.9.6

### 4.9.6 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) und [ZooKeeper-Sandbox](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

### Informationen zur Anwendungsversion

	emr-4.9.6	emr-4.9.5	emr-4.9.4	emr-4.9.3
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.2	1.2.2
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-

	emr-4.9.6	emr-4.9.5	emr-4.9.4	emr-4.9.3
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.14.0	0,14,0	0,14,0	0,14,0
Presto	-	-	-	-
Presto-Sandbox	0.157.1	0.157.1	0.157.1	0.157.1
Spark	1.6.3	1.6.3	1.6.3	1.6.3
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-

	emr-4.9.6	emr-4.9.5	emr-4.9.4	emr-4.9.3
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.6.1	0.6.1	0.6.1	0.6.1
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.9	3.4.9	3.4.9	3.4.9

## Versionshinweise zu 4.9.6

Dies ist eine Patch-Version, um die AWS Signature Version 4-Authentifizierung für Anfragen an Amazon S3 hinzuzufügen. Alle Anwendungen und Komponenten sind dieselben wie in der vorherigen Amazon-EMR-Version.

### Important

In dieser Release-Version verwendet Amazon EMR AWS Signature Version 4 ausschließlich zur Authentifizierung von Anfragen an Amazon S3. Weitere Informationen finden Sie unter [Was ist neu](#).

## 4.9.6 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die

Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als 2.2-amzn-2 aufgeführt.

Komponente	Version	Beschreibung
emr-ddb	4.3.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.2.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.3.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.4.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.17.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den



Komponente	Version	Beschreibung
		Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-2	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-2	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httfs-server	2.7.3-amzn-2	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.7.3-amzn-2	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.7.3-amzn-2	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.

Komponente	Version	Beschreibung
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.2.2	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.2.2	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.2.2	HBase-Befehlszeilen-Client.
hbase-rest-server	1.2.2	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.2.2	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	1.0.0-amzn-9	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	1.0.0-amzn-9	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).

Komponente	Version	Beschreibung
hcatalog-webhcat-server	1.0.0-amzn-9	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	1.0.0-amzn-9	Hive-Befehlszeilen-Client.
hive-metastore-server	1.0.0-amzn-9	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server	1.0.0-amzn-9	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.7.1-amzn-7	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.
oozie-client	4.2.0	Oozie-Befehlszeilen-Client.
oozie-server	4.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.7.0-HBase-1.2	Die Phoenix-Bibliotheken für den Server und den Client

Komponente	Version	Beschreibung
phoenix-query-server	4.7.0-HBase-1.2	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.157.1	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.157.1	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.14.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	1.6.3	Spark-Befehlszeilen-Clients.
spark-history-server	1.6.3	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	1.6.3	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	1.6.3	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.

Komponente	Version	Beschreibung
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.6.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.9	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.9	ZooKeeper Befehlszeilenclient.

#### 4.9.6 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

##### emr-4.9.6-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.

Klassifizierungen	Beschreibung
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.

Klassifizierungen	Beschreibung
hcatalog-webhcat-log4j	Ändert die Werte in der log4j.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j	Ändern Sie die Werte in der Datei 4j.properties von hive-exec-log Hive.
hive-log4j	Ändert die Werte in der hive-log4j.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.

Klassifizierungen	Beschreibung
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.



Klassifizierungen	Beschreibung
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.

Klassifizierungen	Beschreibung
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 4.9.5

### 4.9.5 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) und [ZooKeeper-Sandbox](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

#### Informationen zur Anwendungsversion

	emr-4.9.5	emr-4.9.4	emr-4.9.3	emr-4.9.2
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-

	emr-4.9.5	emr-4.9.4	emr-4.9.3	emr-4.9.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.2	1.2.2
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.14.0	0,14,0	0,14,0	0,14,0
Presto	-	-	-	-
Presto-Sandbox	0.157.1	0.157.1	0.157.1	0.157.1
Spark	1.6.3	1.6.3	1.6.3	1.6.3

	emr-4.9.5	emr-4.9.4	emr-4.9.3	emr-4.9.2
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.6.1	0.6.1	0.6.1	0.6.1
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.9	3.4.9	3.4.9	3.4.9

## Versionshinweise zu 4.9.5

Die folgenden Versionshinweise enthalten Informationen über Amazon EMR Version 4.9.5. Änderungen beziehen sich auf Version 4.9.4.

Erste Version: 29. August 2018

Änderungen, Verbesserungen und behobene Probleme

- HBase
  - Diese Version behebt eine potenzielle Schwachstellen.

## 4.9.5 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon

EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>emr-ddb</code>	4.3.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.2.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.3.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-s3-dist-cp</code>	2.4.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
<code>emrfs</code>	2.17.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>ganglia-monitor</code>	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.

Komponente	Version	Beschreibung
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-2	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-2	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httpfs-server	2.7.3-amzn-2	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.7.3-amzn-2	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider

Komponente	Version	Beschreibung
hadoop-mapred	2.7.3-amzn-2	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.2.2	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.2.2	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.2.2	HBase-Befehlszeilen-Client.
hbase-rest-server	1.2.2	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.2.2	Service, der einen Thrift-Endpunkt für HBase bereitstellt.



Komponente	Version	Beschreibung
hcatalog-client	1.0.0-amzn-9	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	1.0.0-amzn-9	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	1.0.0-amzn-9	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	1.0.0-amzn-9	Hive-Befehlszeilen-Client.
hive-metastore-server	1.0.0-amzn-9	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server	1.0.0-amzn-9	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.7.1-amzn-7	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.
oozie-client	4.2.0	Oozie-Befehlszeilen-Client.

Komponente	Version	Beschreibung
oozie-server	4.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.7.0-HBase-1.2	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.7.0-HBase-1.2	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.157.1	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.157.1	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.14.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	1.6.3	Spark-Befehlszeilen-Clients.
spark-history-server	1.6.3	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	1.6.3	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	1.6.3	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.

Komponente	Version	Beschreibung
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.6.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.9	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.9	ZooKeeper Befehlszeilenclient.

#### 4.9.5 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

##### emr-4.9.5-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.

Klassifizierungen	Beschreibung
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.

Klassifizierungen	Beschreibung
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j	Ändert die Werte in der log4j.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j	Ändern Sie die Werte in der Datei 4j.properties von hive-exec-log Hive.
hive-log4j	Ändert die Werte in der hive-log4j.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.

Klassifizierungen	Beschreibung
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.

Klassifizierungen	Beschreibung
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.



Klassifizierungen	Beschreibung
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 4.9.4

### 4.9.4 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) und [ZooKeeper-Sandbox](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

### Informationen zur Anwendungsversion

	emr-4.9.4	emr-4.9.3	emr-4.9.2	emr-4.9.1
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt

	emr-4.9.4	emr-4.9.3	emr-4.9.2	emr-4.9.1
Scala	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.2	1.2.2
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0

	emr-4.9.4	emr-4.9.3	emr-4.9.2	emr-4.9.1
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.14.0	0,14,0	0,14,0	0,14,0
Presto	-	-	-	-
Presto-Sandbox	0.157.1	0.157.1	0.157.1	0.157.1
Spark	1.6.3	1.6.3	1.6.3	1.6.3
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.6.1	0.6.1	0.6.1	0.6.1
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.9	3.4.9	3.4.9	3.4.9

## Versionshinweise zu 4.9.4

Die folgenden Versionshinweise enthalten Informationen über Amazon EMR Version 4.9.4. Änderungen beziehen sich auf Version 4.9.3.

Erste Version: 29. März 2018

## Änderungen, Verbesserungen und behobene Probleme

- Der Amazon-Linux-Kernel des standardmäßigen Amazon-Linux-AMI für Amazon EMR wurde aktualisiert, um potenzielle Schwachstellen zu schließen.

### 4.9.4 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>emr-ddb</code>	4.3.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.2.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.3.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-s3-dist-cp</code>	2.4.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.

Komponente	Version	Beschreibung
emrfs	2.17.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-2	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-2	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.

Komponente	Version	Beschreibung
hadoop-https-server	2.7.3-amzn-2	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.7.3-amzn-2	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.7.3-amzn-2	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resource-manager	2.7.3-amzn-2	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.2.2	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.2.2	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.

Komponente	Version	Beschreibung
hbase-client	1.2.2	HBase-Befehlszeilen-Client.
hbase-rest-server	1.2.2	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.2.2	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	1.0.0-amzn-9	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	1.0.0-amzn-9	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	1.0.0-amzn-9	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	1.0.0-amzn-9	Hive-Befehlszeilen-Client.
hive-metastore-server	1.0.0-amzn-9	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server	1.0.0-amzn-9	Service zur Annahme von Hive-Abfragen als Webanfragen.

Komponente	Version	Beschreibung
hue-server	3.7.1-amzn-7	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.
oozie-client	4.2.0	Oozie-Befehlszeilen-Client.
oozie-server	4.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.7.0-HBase-1.2	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.7.0-HBase-1.2	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.157.1	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.157.1	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.14.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	1.6.3	Spark-Befehlszeilen-Clients.



Komponente	Version	Beschreibung
spark-history-server	1.6.3	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	1.6.3	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	1.6.3	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.6.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.9	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.9	ZooKeeper Befehlszeilenclient.

## 4.9.4 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

### emr-4.9.4-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der <code>hbase-log4j.properties</code> -Datei in HBase.
hbase-metrics	Ändert die Werte in der <code>hadoop-metrics2-hbase.properties</code> -Datei in HBase.
hbase-policy	Ändert die Werte in der <code>hbase-policy.xml</code> -Datei in HBase.

Klassifizierungen	Beschreibung
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j	Ändert die Werte in der log4j.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j	Ändern Sie die Werte in der Datei 4j.properties von hive-exec-log Hive.
hive-log4j	Ändert die Werte in der hive-log4j.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.

Klassifizierungen	Beschreibung
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.

Klassifizierungen	Beschreibung
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.

Klassifizierungen	Beschreibung
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 4.9.3

### 4.9.3 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) und [ZooKeeper-Sandbox](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

## Informationen zur Anwendungsversion

	emr-4.9.3	emr-4.9.2	emr-4.9.1	emr-4.8.5
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.2	1.2.2
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-



	emr-4.9.3	emr-4.9.2	emr-4.9.1	emr-4.8.5
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.14.0	0,14,0	0,14,0	0,14,0
Presto	-	-	-	-
Presto-Sandbox	0.157.1	0.157.1	0.157.1	0.157.1
Spark	1.6.3	1.6.3	1.6.3	1.6.3
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.6.1	0.6.1	0.6.1	0.6.1
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.9	3.4.9	3.4.9	3.4.9

## Versionshinweise zu 4.9.3

Die folgenden Versionshinweise enthalten Informationen über Amazon EMR Version 4.9.3. Änderungen beziehen sich auf Amazon EMR Version 4.9.2.

Erste Version: 22. Januar 2018

Änderungen, Verbesserungen und behobene Probleme

- Der Amazon-Linux-Kernel des standardmäßigen Amazon-Linux-AMI für Amazon EMR wurde aktualisiert, um Schwachstellen im Zusammenhang mit spekulativer Ausführung zu beheben (CVE-2017-5715, CVE-2017-5753 und CVE-2017-5754). Weitere Informationen finden Sie unter <https://aws.amazon.com/security/security-bulletins/AWS-2018-013/>.

### 4.9.3 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>emr-ddb</code>	4.3.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.2.0	Praktische Bibliotheken für das Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-kinesis	3.3.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.4.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.17.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-2	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFS-Service auf Knotenebene zum Speichern von Blöcken.

Komponente	Version	Beschreibung
hadoop-hdfs-library	2.7.3-amzn-2	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httpfs-server	2.7.3-amzn-2	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.7.3-amzn-2	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.7.3-amzn-2	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.

Komponente	Version	Beschreibung
hbase-hmaster	1.2.2	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.2.2	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.2.2	HBase-Befehlszeilen-Client.
hbase-rest-server	1.2.2	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.2.2	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	1.0.0-amzn-9	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	1.0.0-amzn-9	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	1.0.0-amzn-9	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	1.0.0-amzn-9	Hive-Befehlszeilen-Client.

Komponente	Version	Beschreibung
hive-metastore-server	1.0.0-amzn-9	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server	1.0.0-amzn-9	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.7.1-amzn-7	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.
oozie-client	4.2.0	Oozie-Befehlszeilen-Client.
oozie-server	4.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.7.0-HBase-1.2	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.7.0-HBase-1.2	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.

Komponente	Version	Beschreibung
presto-coordinator	0.157.1	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.157.1	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.14.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	1.6.3	Spark-Befehlszeilen-Clients.
spark-history-server	1.6.3	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	1.6.3	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	1.6.3	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.6.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.

Komponente	Version	Beschreibung
zookeeper-server	3.4.9	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.9	ZooKeeper Befehlszeilenclient.

### 4.9.3 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

#### emr-4.9.3-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.



Klassifizierungen	Beschreibung
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j	Ändert die Werte in der log4j.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.

Klassifizierungen	Beschreibung
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j	Ändern Sie die Werte in der Datei 4j.properties von hive-exec-log Hive.
hive-log4j	Ändert die Werte in der hive-log4j.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.

Klassifizierungen	Beschreibung
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.

Klassifizierungen	Beschreibung
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 4.9.2

### 4.9.2 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) und [ZooKeeper-Sandbox](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

#### Informationen zur Anwendungsversion

	emr-4.9.2	emr-4.9.1	emr-4.8.5	emr-4.8.4
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.2	1.2.2
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3

	emr-4.9.2	emr-4.9.1	emr-4.8.5	emr-4.8.4
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.14.0	0,14,0	0,14,0	0,14,0
Presto	-	-	-	-
Presto-Sandbox	0.157.1	0.157.1	0.157.1	0.157.1
Spark	1.6.3	1.6.3	1.6.3	1.6.3
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4

	emr-4.9.2	emr-4.9.1	emr-4.8.5	emr-4.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.6.1	0.6.1	0.6.1	0.6.1
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.9	3.4.9	3.4.9	3.4.9

## Versionshinweise zu 4.9.2

Die folgenden Versionshinweise enthalten Informationen über Amazon EMR Version 4.9.2. Änderungen beziehen sich auf Amazon EMR Version 4.9.1.

Veröffentlichungsdatum: 13. Juli 2017

Kleinere Änderungen, Fehlerbehebungen und Verbesserungen wurden in dieser Version vorgenommen.

### 4.9.2 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die



Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als 2.2-amzn-2 aufgeführt.

Komponente	Version	Beschreibung
emr-ddb	4.3.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.2.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.3.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.4.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.17.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den

Komponente	Version	Beschreibung
		Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-2	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-2	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httfs-server	2.7.3-amzn-2	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.7.3-amzn-2	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.7.3-amzn-2	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.

Komponente	Version	Beschreibung
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.2.2	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.2.2	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.2.2	HBase-Befehlszeilen-Client.
hbase-rest-server	1.2.2	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.2.2	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	1.0.0-amzn-9	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	1.0.0-amzn-9	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).

Komponente	Version	Beschreibung
hcatalog-webhcat-server	1.0.0-amzn-9	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	1.0.0-amzn-9	Hive-Befehlszeilen-Client.
hive-metastore-server	1.0.0-amzn-9	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server	1.0.0-amzn-9	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.7.1-amzn-7	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.
oozie-client	4.2.0	Oozie-Befehlszeilen-Client.
oozie-server	4.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.7.0-HBase-1.2	Die Phoenix-Bibliotheken für den Server und den Client

Komponente	Version	Beschreibung
phoenix-query-server	4.7.0-HBase-1.2	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.157.1	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.157.1	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.14.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	1.6.3	Spark-Befehlszeilen-Clients.
spark-history-server	1.6.3	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	1.6.3	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	1.6.3	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.

Komponente	Version	Beschreibung
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.6.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.9	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.9	ZooKeeper Befehlszeilenclient.

## 4.9.2 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

### emr-4.9.2-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.

Klassifizierungen	Beschreibung
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.

Klassifizierungen	Beschreibung
hcatalog-webhcat-log4j	Ändert die Werte in der log4j.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j	Ändern Sie die Werte in der Datei 4j.properties von hive-exec-log Hive.
hive-log4j	Ändert die Werte in der hive-log4j.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.



Klassifizierungen	Beschreibung
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.

Klassifizierungen	Beschreibung
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 4.9.1

### 4.9.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) und [ZooKeeper-Sandbox](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

#### Informationen zur Anwendungsversion

	emr-4.9.1	emr-4.8.5	emr-4.8.4	emr-4.8.3
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-

	emr-4.9.1	emr-4.8.5	emr-4.8.4	emr-4.8.3
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.2	1.2.2
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.14.0	0,14,0	0,14,0	0,14,0
Presto	-	-	-	-
Presto-Sandbox	0.157.1	0.157.1	0.157.1	0.157.1
Spark	1.6.3	1.6.3	1.6.3	1.6.3

	emr-4.9.1	emr-4.8.5	emr-4.8.4	emr-4.8.3
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.6.1	0.6.1	0.6.1	0.6.1
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.9	3.4.9	3.4.9	3.4.9

## Versionshinweise zu 4.9.1

Die folgenden Versionshinweise enthalten Informationen über Amazon EMR Version 4.9.1. Änderungen beziehen sich auf Amazon EMR Version 4.8.4.

Veröffentlichungsdatum: 10. April 2017

Bekannte Probleme aus den früheren Versionen, die behoben wurden

- Rückportierungen von [HIVE-9976](#) und [HIVE-10106](#)
- Es wurde ein Problem in YARN behoben, bei dem eine große Anzahl von Knoten (mehr als 2.000) und Containern (mehr als 5.000) einen out-of-memory Fehler verursachte, zum Beispiel: "Exception in thread main java.lang.OutOfMemoryError".

## Änderungen und Verbesserungen

- Amazon-EMR-Versionen basieren jetzt auf Amazon Linux 2017.03. Weitere Informationen finden Sie unter <https://aws.amazon.com/amazon-linux-ami/2017.03-release-notes/>.
- Python 2.6 wurde aus dem Amazon-EMR-Basis-Linux-Image entfernt. Sie können Python 2.6 bei Bedarf manuell installieren.

### 4.9.1-Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>emr-ddb</code>	4.2.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.2.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.3.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-s3-dist-cp	2.4.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.15.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-2	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-2	HDFS-Client und -Bibliothek für die Befehlszeile



Komponente	Version	Beschreibung
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httpfs-server	2.7.3-amzn-2	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.7.3-amzn-2	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.7.3-amzn-2	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.

Komponente	Version	Beschreibung
hbase-hmaster	1.2.2	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.2.2	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.2.2	HBase-Befehlszeilen-Client.
hbase-rest-server	1.2.2	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.2.2	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	1.0.0-amzn-9	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	1.0.0-amzn-9	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	1.0.0-amzn-9	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	1.0.0-amzn-9	Hive-Befehlszeilen-Client.

Komponente	Version	Beschreibung
hive-metastore-server	1.0.0-amzn-9	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server	1.0.0-amzn-9	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.7.1-amzn-7	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.
oozie-client	4.2.0	Oozie-Befehlszeilen-Client.
oozie-server	4.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.7.0-HBase-1.2	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.7.0-HBase-1.2	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.

Komponente	Version	Beschreibung
presto-coordinator	0.157.1	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.157.1	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.14.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	1.6.3	Spark-Befehlszeilen-Clients.
spark-history-server	1.6.3	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	1.6.3	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	1.6.3	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.6.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.

Komponente	Version	Beschreibung
zookeeper-server	3.4.9	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.9	ZooKeeper Befehlszeilenclient.

#### 4.9.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

##### emr-4.9.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.

Klassifizierungen	Beschreibung
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j	Ändert die Werte in der log4j.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.

Klassifizierungen	Beschreibung
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j	Ändern Sie die Werte in der Datei 4j.properties von hive-exec-log Hive.
hive-log4j	Ändert die Werte in der hive-log4j.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.

Klassifizierungen	Beschreibung
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.



Klassifizierungen	Beschreibung
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.

Klassifizierungen	Beschreibung
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 4.8.5

### 4.8.5 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) und [ZooKeeper-Sandbox](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

#### Informationen zur Anwendungsversion

	emr-4.8.5	emr-4.8.4	emr-4.8.3	emr-4.8.2
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.2	1.2.2
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3

	emr-4.8.5	emr-4.8.4	emr-4.8.3	emr-4.8.2
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.14.0	0,14,0	0,14,0	0,14,0
Presto	-	-	-	-
Presto-Sandbox	0.157.1	0.157.1	0.157.1	0.152.3
Spark	1.6.3	1.6.3	1.6.3	1.6.2
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4

	emr-4.8.5	emr-4.8.4	emr-4.8.3	emr-4.8.2
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.6.1	0.6.1	0.6.1	0.6.1
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.9	3.4.9	3.4.9	3.4.8

## Versionshinweise zu 4.8.5

Dies ist eine Patch-Version, um die AWS Signature Version 4-Authentifizierung für Anfragen an Amazon S3 hinzuzufügen. Alle Anwendungen und Komponenten sind dieselben wie in der vorherigen Amazon-EMR-Version.

### Important

In dieser Release-Version verwendet Amazon EMR AWS Signature Version 4 ausschließlich zur Authentifizierung von Anfragen an Amazon S3. Weitere Informationen finden Sie unter [Was ist neu](#).

## 4.8.5-Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-*

amzn-*EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen myapp-component der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als 2.2-amzn-2 aufgeführt.

Komponente	Version	Beschreibung
emr-ddb	4.2.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.2.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.2.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.4.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.14.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den

Komponente	Version	Beschreibung
		Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-1	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-1	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.7.3-amzn-1	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httfs-server	2.7.3-amzn-1	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.7.3-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.7.3-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-1	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.

Komponente	Version	Beschreibung
hadoop-yarn-resourcemanager	2.7.3-amzn-1	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-1	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.2.2	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.2.2	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.2.2	HBase-Befehlszeilen-Client.
hbase-rest-server	1.2.2	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.2.2	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	1.0.0-amzn-8	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	1.0.0-amzn-8	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).



Komponente	Version	Beschreibung
hcatalog-webhcat-server	1.0.0-amzn-8	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	1.0.0-amzn-8	Hive-Befehlszeilen-Client.
hive-metastore-server	1.0.0-amzn-8	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server	1.0.0-amzn-8	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.7.1-amzn-7	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.
oozie-client	4.2.0	Oozie-Befehlszeilen-Client.
oozie-server	4.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.7.0-HBase-1.2	Die Phoenix-Bibliotheken für den Server und den Client

Komponente	Version	Beschreibung
phoenix-query-server	4.7.0-HBase-1.2	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.157.1	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.157.1	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.14.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	1.6.3	Spark-Befehlszeilen-Clients.
spark-history-server	1.6.3	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	1.6.3	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	1.6.3	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.

Komponente	Version	Beschreibung
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.6.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.9	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.9	ZooKeeper Befehlszeilenclient.

## 4.8.5 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

### emr-4.8.5-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.

Klassifizierungen	Beschreibung
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.

Klassifizierungen	Beschreibung
hcatalog-webhcat-log4j	Ändert die Werte in der log4j.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j	Ändern Sie die Werte in der Datei 4j.properties von hive-exec-log Hive.
hive-log4j	Ändert die Werte in der hive-log4j.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.

Klassifizierungen	Beschreibung
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.

Klassifizierungen	Beschreibung
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.



## Amazon-EMR-Version 4.8.4

### 4.8.4 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) und [ZooKeeper-Sandbox](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

#### Informationen zur Anwendungsversion

	emr-4.8.4	emr-4.8.3	emr-4.8.2	emr-4.8.1
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-

	emr-4.8.4	emr-4.8.3	emr-4.8.2	emr-4.8.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.2	1.2.2
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.2
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.14.0	0,14,0	0,14,0	0,14,0
Presto	-	-	-	-
Presto-Sandbox	0.157.1	0.157.1	0.152.3	0.151
Spark	1.6.3	1.6.3	1.6.2	1.6.2

	emr-4.8.4	emr-4.8.3	emr-4.8.2	emr-4.8.1
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.6.1	0.6.1	0.6.1	0.6.1
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.9	3.4.9	3.4.8	3.4.8

## Versionshinweise zu 4.8.4

Die folgenden Versionshinweise enthalten Informationen über Amazon EMR Version 4.8.4. Änderungen beziehen sich auf Amazon EMR Version 4.8.3.

Veröffentlichungsdatum: 7. Februar 2017

Kleinere Änderungen, Fehlerbehebungen und Verbesserungen wurden in dieser Version vorgenommen.

### 4.8.4 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit emr oder aws. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version

sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-amzn-*EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen myapp-component der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als 2.2-amzn-2 aufgeführt.

Komponente	Version	Beschreibung
emr-ddb	4.2.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.2.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.2.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.4.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.14.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken

Komponente	Version	Beschreibung
		aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-1	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-1	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.7.3-amzn-1	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httpfs-server	2.7.3-amzn-1	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.7.3-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.7.3-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.

Komponente	Version	Beschreibung
hadoop-yarn-nodemanager	2.7.3-amzn-1	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.3-amzn-1	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-1	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.2.2	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.2.2	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.2.2	HBase-Befehlszeilen-Client.
hbase-rest-server	1.2.2	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.2.2	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	1.0.0-amzn-8	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.

Komponente	Version	Beschreibung
hcatalog-server	1.0.0-amzn-8	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	1.0.0-amzn-8	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	1.0.0-amzn-8	Hive-Befehlszeilen-Client.
hive-metastore-server	1.0.0-amzn-8	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server	1.0.0-amzn-8	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.7.1-amzn-7	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.54+	MySQL-Datenbankserver.
oozie-client	4.2.0	Oozie-Befehlszeilen-Client.
oozie-server	4.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.

Komponente	Version	Beschreibung
phoenix-library	4.7.0-HBase-1.2	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.7.0-HBase-1.2	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.157.1	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.157.1	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.14.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	1.6.3	Spark-Befehlszeilen-Clients.
spark-history-server	1.6.3	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	1.6.3	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	1.6.3	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.



Komponente	Version	Beschreibung
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.25+	Apache HTTP-Server.
zeppelin-server	0.6.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.9	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.9	ZooKeeper Befehlszeilenclient.

#### 4.8.4 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

##### emr-4.8.4-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.

Klassifizierungen	Beschreibung
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der .xml-Datei von HCatalog. proto-hive-site

Klassifizierungen	Beschreibung
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j	Ändert die Werte in der log4j.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j	Ändern Sie die Werte in der Datei 4j.properties von hive-exec-log Hive.
hive-log4j	Ändert die Werte in der hive-log4j.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.

Klassifizierungen	Beschreibung
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.

Klassifizierungen	Beschreibung
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 4.8.3

### 4.8.3 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) und [ZooKeeper-Sandbox](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

#### Informationen zur Anwendungsversion

	emr-4.8.3	emr-4.8.2	emr-4.8.1	emr-4.8.0
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-

	emr-4.8.3	emr-4.8.2	emr-4.8.1	emr-4.8.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.2	1.2.2
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.3	2.7.3	2.7.2	2.7.2
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.14.0	0,14,0	0,14,0	0,14,0
Presto	-	-	-	-
Presto-Sandbox	0.157.1	0.152.3	0.151	0.151
Spark	1.6.3	1.6.2	1.6.2	1.6.2



	emr-4.8.3	emr-4.8.2	emr-4.8.1	emr-4.8.0
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.6.1	0.6.1	0.6.1	0.6.1
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.9	3.4.8	3.4.8	3.4.8

## Versionshinweise zu 4.8.3

Die folgenden Versionshinweise enthalten Informationen über Amazon EMR Version 4.8.3. Änderungen beziehen sich auf Amazon EMR Version 4.8.2.

Veröffentlichungsdatum: 29. Dezember 2016

### Upgrades

- Upgrade auf Presto 0.157.1. Weitere Informationen finden Sie in den [Presto-Versionshinweisen](#) in der Presto-Dokumentation.
- Upgrade auf Spark 1.6.3. Weitere Informationen finden Sie in den [Spark-Versionshinweisen](#) in der Apache Spark-Dokumentation.
- Auf ZooKeeper 3.4.9 aktualisiert. Weitere Informationen finden Sie in den [ZooKeeper Versionshinweisen](#) in der ZooKeeper Apache-Dokumentation.

## Änderungen und Verbesserungen

- Unterstützung hinzugefügt für den Amazon-EC2-Instance-Typ m4.16xlarge in Amazon-EMR-Version 4.8.3 und höher, ausgenommen 5.0.0, 5.0.3 und 5.2.0.
- Amazon-EMR-Versionen basieren jetzt auf Amazon Linux 2016.09. Weitere Informationen finden Sie unter <https://aws.amazon.com/amazon-linux-ami/2016.09-release-notes/>.

### Bekannte Probleme aus den früheren Versionen, die behoben wurden

- Es wurde ein Problem in Hadoop behoben, bei dem der ReplicationMonitor Thread aufgrund eines Wettlaufs zwischen Replikation und Löschung derselben Datei in einem großen Cluster für lange Zeit hängen bleiben konnte.
- Es wurde ein Problem behoben, bei dem ControlledJob #toString mit einer Nullzeiger-Ausnahme (NPE) fehlschlug, wenn der Jobstatus nicht erfolgreich aktualisiert wurde.

## 4.8.3-Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit emr oder aws. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-amzn-*EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen myapp-component der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als 2.2-amzn-2 aufgeführt.

Komponente	Version	Beschreibung
emr-ddb	4.2.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-goodies	2.2.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.2.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.4.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.13.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".

Komponente	Version	Beschreibung
hadoop-hdfs-datanode	2.7.3-amzn-1	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-1	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.7.3-amzn-1	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httpfs-server	2.7.3-amzn-1	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.7.3-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.7.3-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.3-amzn-1	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resource-manager	2.7.3-amzn-1	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.

Komponente	Version	Beschreibung
hadoop-yarn-timeline-server	2.7.3-amzn-1	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.2.2	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.2.2	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.2.2	HBase-Befehlszeilen-Client.
hbase-rest-server	1.2.2	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.2.2	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	1.0.0-amzn-8	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	1.0.0-amzn-8	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	1.0.0-amzn-8	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.

Komponente	Version	Beschreibung
hive-client	1.0.0-amzn-8	Hive-Befehlszeilen-Client.
hive-metastore-server	1.0.0-amzn-8	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server	1.0.0-amzn-8	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.7.1-amzn-7	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.52	MySQL-Datenbankserver.
oozie-client	4.2.0	Oozie-Befehlszeilen-Client.
oozie-server	4.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.7.0-HBase-1.2	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.7.0-HBase-1.2	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.

Komponente	Version	Beschreibung
presto-coordinator	0.157.1	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.157.1	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.14.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	1.6.3	Spark-Befehlszeilen-Clients.
spark-history-server	1.6.3	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	1.6.3	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	1.6.3	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.23	Apache HTTP-Server.
zeppelin-server	0.6.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.

Komponente	Version	Beschreibung
zookeeper-server	3.4.9	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.9	ZooKeeper Befehlszeilenclient.

### 4.8.3 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

#### emr-4.8.3-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.



Klassifizierungen	Beschreibung
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der .xml-Datei von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j	Ändert die Werte in der log4j.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.

Klassifizierungen	Beschreibung
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j	Ändern Sie die Werte in der Datei 4j.properties von hive-exec-log Hive.
hive-log4j	Ändert die Werte in der hive-log4j.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.

Klassifizierungen	Beschreibung
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.

Klassifizierungen	Beschreibung
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 4.8.2

### 4.8.2 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) und [ZooKeeper-Sandbox](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

#### Informationen zur Anwendungsversion

	emr-4.8.2	emr-4.8.1	emr-4.8.0	emr-4.7.4
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.2	1.2.1
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.3	2.7.2	2.7.2	2.7.2

	emr-4.8.2	emr-4.8.1	emr-4.8.0	emr-4.7.4
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.14.0	0,14,0	0,14,0	0,14,0
Presto	-	-	-	-
Presto-Sandbox	0.152.3	0.151	0.151	0.148
Spark	1.6.2	1.6.2	1.6.2	1.6.2
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.3

	emr-4.8.2	emr-4.8.1	emr-4.8.0	emr-4.7.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.6.1	0.6.1	0.6.1	0.5.6
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.8	3.4.8	3.4.8	3.4.8

## 4.8.2 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon EMR Version 4.8.2. Änderungen beziehen sich auf Amazon EMR Version 4.8.0.

Veröffentlichungsdatum: 24. Oktober 2016

### Upgrades

- Upgrade auf Hadoop 2.7.3
- Es wurde ein Upgrade auf Presto 0.152.3 durchgeführt, das Support für die Presto-Webschnittstelle umfasst. Sie können auf die Presto-Webschnittstelle des Presto-Koordinators über Port 8889 zugreifen. Weitere Informationen über die Presto-Webschnittstelle finden Sie unter [Web Interface](#) in der Presto-Dokumentation.
- Amazon-EMR-Versionen basieren jetzt auf Amazon Linux 2016.09. Weitere Informationen finden Sie unter <https://aws.amazon.com/amazon-linux-ami/2016.09-release-notes/>.

## 4.8.2-Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit emr oder aws. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version



sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-amzn-*EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen myapp-component der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als 2.2-amzn-2 aufgeführt.

Komponente	Version	Beschreibung
emr-ddb	4.1.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.1.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.2.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.4.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.10.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken

Komponente	Version	Beschreibung
		aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.3-amzn-0	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.3-amzn-0	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.3-amzn-0	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.7.3-amzn-0	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httpfs-server	2.7.3-amzn-0	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.7.3-amzn-0	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.7.3-amzn-0	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.

Komponente	Version	Beschreibung
hadoop-yarn-nodemanager	2.7.3-amzn-0	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.3-amzn-0	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.3-amzn-0	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.2.2	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.2.2	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.2.2	HBase-Befehlszeilen-Client.
hbase-rest-server	1.2.2	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.2.2	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	1.0.0-amzn-7	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.

Komponente	Version	Beschreibung
hcatalog-server	1.0.0-amzn-7	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	1.0.0-amzn-7	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	1.0.0-amzn-7	Hive-Befehlszeilen-Client.
hive-metastore-server	1.0.0-amzn-7	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server	1.0.0-amzn-7	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.7.1-amzn-7	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.52	MySQL-Datenbankserver.
oozie-client	4.2.0	Oozie-Befehlszeilen-Client.
oozie-server	4.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.

Komponente	Version	Beschreibung
phoenix-library	4.7.0-HBase-1.2	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.7.0-HBase-1.2	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.152.3	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.152.3	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.14.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	1.6.2	Spark-Befehlszeilen-Clients.
spark-history-server	1.6.2	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	1.6.2	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	1.6.2	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.

Komponente	Version	Beschreibung
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.23	Apache HTTP-Server.
zeppelin-server	0.6.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.8	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.8	ZooKeeper Befehlszeilenclient.

## 4.8.2 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

### emr-4.8.2-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.

Klassifizierungen	Beschreibung
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der .xml-Datei von HCatalog. proto-hive-site

Klassifizierungen	Beschreibung
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j	Ändert die Werte in der log4j.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j	Ändern Sie die Werte in der Datei 4j.properties von hive-exec-log Hive.
hive-log4j	Ändert die Werte in der hive-log4j.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.



Klassifizierungen	Beschreibung
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.

Klassifizierungen	Beschreibung
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon EMR Version 4.8.1

### 4.8.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) und [ZooKeeper-Sandbox](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

#### Informationen zur Anwendungsversion

	emr-4.8.1	emr-4.8.0	emr-4.7.4	emr-4.7.3
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-

	emr-4.8.1	emr-4.8.0	emr-4.7.4	emr-4.7.3
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.1	1.2.1
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.2	2.7.2	2.7.2	2.7.2
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.14.0	0,14,0	0,14,0	0,14,0
Presto	-	-	-	-
Presto-Sandbox	0.151	0.151	0.148	0.148
Spark	1.6.2	1.6.2	1.6.2	1.6.2

	emr-4.8.1	emr-4.8.0	emr-4.7.4	emr-4.7.3
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.3	0.8.3
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.6.1	0.6.1	0.5.6	0.5.6
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.8	3.4.8	3.4.8	3.4.8

## Versionshinweise zu 4.8.1

### 4.8.1-Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die

Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als 2.2-amzn-2 aufgeführt.

Komponente	Version	Beschreibung
emr-ddb	3.2.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.1.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.2.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.4.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.9.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den

Komponente	Version	Beschreibung
		Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.2-amzn-4	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.2-amzn-4	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.2-amzn-4	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.7.2-amzn-4	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httfs-server	2.7.2-amzn-4	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.7.2-amzn-4	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.7.2-amzn-4	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.2-amzn-4	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.



Komponente	Version	Beschreibung
hadoop-yarn-resourcemanager	2.7.2-amzn-4	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.2-amzn-4	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.2.2	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.2.2	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.2.2	HBase-Befehlszeilen-Client.
hbase-rest-server	1.2.2	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.2.2	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	1.0.0-amzn-7	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	1.0.0-amzn-7	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).

Komponente	Version	Beschreibung
hcatalog-webhcat-server	1.0.0-amzn-7	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	1.0.0-amzn-7	Hive-Befehlszeilen-Client.
hive-metastore-server	1.0.0-amzn-7	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server	1.0.0-amzn-7	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.7.1-amzn-7	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.51	MySQL-Datenbankserver.
oozie-client	4.2.0	Oozie-Befehlszeilen-Client.
oozie-server	4.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.7.0-HBase-1.2	Die Phoenix-Bibliotheken für den Server und den Client

Komponente	Version	Beschreibung
phoenix-query-server	4.7.0-HBase-1.2	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.151	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.151	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.14.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	1.6.2	Spark-Befehlszeilen-Clients.
spark-history-server	1.6.2	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	1.6.2	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	1.6.2	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.

Komponente	Version	Beschreibung
webserver	2.4.23	Apache HTTP-Server.
zeppelin-server	0.6.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.8	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.8	ZooKeeper Befehlszeilenclient.

#### 4.8.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

##### emr-4.8.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.

Klassifizierungen	Beschreibung
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.

Klassifizierungen	Beschreibung
hcatalog-webhcat-log4j	Ändert die Werte in der log4j.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j	Ändern Sie die Werte in der Datei 4j.properties von hive-exec-log Hive.
hive-log4j	Ändert die Werte in der hive-log4j.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.

Klassifizierungen	Beschreibung
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.



Klassifizierungen	Beschreibung
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 4.8.0

### 4.8.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) und [ZooKeeper-Sandbox](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

#### Informationen zur Anwendungsversion

	emr-4.8.0	emr-4.7.4	emr-4.7.3	emr-4.7.2
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-

	emr-4.8.0	emr-4.7.4	emr-4.7.3	emr-4.7.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.1	1.2.1	1.2.1
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.2	2.7.2	2.7.2	2.7.2
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.14.0	0,14,0	0,14,0	0,14,0
Presto	-	-	-	-
Presto-Sandbox	0.151	0.148	0.148	0.148
Spark	1.6.2	1.6.2	1.6.2	1.6.2

	emr-4.8.0	emr-4.7.4	emr-4.7.3	emr-4.7.2
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.3	0.8.3	0.8.3
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.6.1	0.5.6	0.5.6	0.5.6
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.8	3.4.8	3.4.8	3.4.8

## Versionshinweise zu 4.8.0

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 4.8.0. Änderungen beziehen sich auf Amazon-EMR-Version 4.7.2.

Veröffentlichungsdatum: 7. September 2016

### Upgrades

- Upgrade auf Hbase 1.2.2
- Upgrade auf Presto-Sandbox 0.151
- Upgrade auf Tez 0.8.4
- Upgrade auf Zeppelin-Sandbox 0.6.1

## Änderungen und Verbesserungen

- Es wurde ein Problem in YARN behoben, ApplicationMaster bei dem versucht wurde, Container zu bereinigen, die nicht mehr existieren, weil ihre Instanzen beendet wurden.
- Die hive-server2-URL für Hive2-Aktionen in den Oozie-Beispielen wurde korrigiert.
- Zusätzliche Presto-Kataloge werden jetzt unterstützt.
- Patches wurden rückportiert: [HIVE-8948](#), [HIVE-12679](#), [HIVE-13405](#), [PHOENIX-3116](#), [HADOOP-12689](#)
- Es werden jetzt Sicherheitskonfigurationen unterstützt, mit deren Hilfe Sie die Verschlüsselungsoptionen einfacher erstellen und anwenden können. Weitere Informationen finden Sie unter [Datenverschlüsselung](#).

### 4.8.0-Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>emr-ddb</code>	3.2.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.1.0	Praktische Bibliotheken für das Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-kinesis	3.2.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.4.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.9.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.2-amzn-4	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.2-amzn-4	HDFS-Service auf Knotenebene zum Speichern von Blöcken.

Komponente	Version	Beschreibung
hadoop-hdfs-library	2.7.2-amzn-4	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.7.2-amzn-4	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httfs-server	2.7.2-amzn-4	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.7.2-amzn-4	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.7.2-amzn-4	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.2-amzn-4	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.2-amzn-4	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.2-amzn-4	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.

Komponente	Version	Beschreibung
hbase-hmaster	1.2.2	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.2.2	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.2.2	HBase-Befehlszeilen-Client.
hbase-rest-server	1.2.2	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.2.2	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	1.0.0-amzn-7	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	1.0.0-amzn-7	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	1.0.0-amzn-7	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	1.0.0-amzn-7	Hive-Befehlszeilen-Client.



Komponente	Version	Beschreibung
hive-metastore-server	1.0.0-amzn-7	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server	1.0.0-amzn-7	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.7.1-amzn-7	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.51	MySQL-Datenbankserver.
oozie-client	4.2.0	Oozie-Befehlszeilen-Client.
oozie-server	4.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.7.0-HBase-1.2	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.7.0-HBase-1.2	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.

Komponente	Version	Beschreibung
presto-coordinator	0.151	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.151	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.14.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	1.6.2	Spark-Befehlszeilen-Clients.
spark-history-server	1.6.2	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	1.6.2	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	1.6.2	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.4	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.23	Apache HTTP-Server.
zeppelin-server	0.6.1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.

Komponente	Version	Beschreibung
zookeeper-server	3.4.8	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.8	ZooKeeper Befehlszeilenclient.

## 4.8.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

### emr-4.8.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.

Klassifizierungen	Beschreibung
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der .xml-Datei von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j	Ändert die Werte in der log4j.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.

Klassifizierungen	Beschreibung
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j	Ändern Sie die Werte in der Datei 4j.properties von hive-exec-log Hive.
hive-log4j	Ändert die Werte in der hive-log4j.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hiveserver2-site	Ändert die Werte in der hiveserver2-site.xml-Datei von Server2 in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.

Klassifizierungen	Beschreibung
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-connector-blackhole	Ändert die Werte in der blackhole.properties-Datei in Presto.
presto-connector-cassandra	Ändert die Werte in der cassandra.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
presto-connector-jmx	Ändert die Werte in der jmx.properties-Datei in Presto.
presto-connector-kafka	Ändert die Werte in der kafka.properties-Datei in Presto.
presto-connector-localfile	Ändert die Werte in der localfile.properties-Datei in Presto.
presto-connector-mongodb	Ändert die Werte in der mongodb.properties-Datei in Presto.
presto-connector-mysql	Ändert die Werte in der mysql.properties-Datei in Presto.
presto-connector-postgresql	Ändert die Werte in der postgresql.properties-Datei in Presto.
presto-connector-raptor	Ändert die Werte in der raptor.properties-Datei in Presto.
presto-connector-redis	Ändert die Werte in der redis.properties-Datei in Presto.
presto-connector-tpch	Ändert die Werte in der tpch.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.

Klassifizierungen	Beschreibung
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 4.7.4

### 4.7.4 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) und [ZooKeeper-Sandbox](#).



In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

#### Informationen zur Anwendungsversion

	emr-4.7.4	emr-4.7.3	emr-4.7.2	emr-4.7.1
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.1	1.2.1	1.2.1	1.2.1
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.2	2.7.2	2.7.2	2.7.2

	emr-4.7.4	emr-4.7.3	emr-4.7.2	emr-4.7.1
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.0
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.14.0	0,14,0	0,14,0	0,14,0
Presto	-	-	-	-
Presto-Sandbox	0.148	0.148	0.148	0.147
Spark	1.6.2	1.6.2	1.6.2	1.6.1
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.3	0.8.3	0.8.3	0.8.3

	emr-4.7.4	emr-4.7.3	emr-4.7.2	emr-4.7.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.5.6	0.5.6	0.5.6	0.5.6
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.8	3.4.8	3.4.8	3.4.8

## Versionshinweise zu 4.7.4

Dies ist eine Patch-Version, um die AWS Signature Version 4-Authentifizierung für Anfragen an Amazon S3 hinzuzufügen. Alle Anwendungen und Komponenten sind dieselben wie in der vorherigen Amazon-EMR-Version.

### Important

In dieser Release-Version verwendet Amazon EMR AWS Signature Version 4 ausschließlich zur Authentifizierung von Anfragen an Amazon S3. Weitere Informationen finden Sie unter [Was ist neu](#).

## 4.7.4 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-*

amzn-*EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen myapp-component der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als 2.2-amzn-2 aufgeführt.

Komponente	Version	Beschreibung
emr-ddb	3.2.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.1.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.2.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.4.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.8.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den

Komponente	Version	Beschreibung
		Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.2-amzn-3	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.2-amzn-3	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.2-amzn-3	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.7.2-amzn-3	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httfs-server	2.7.2-amzn-3	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.7.2-amzn-3	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.7.2-amzn-3	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.2-amzn-3	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.

Komponente	Version	Beschreibung
hadoop-yarn-resourcemanager	2.7.2-amzn-3	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.2-amzn-3	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.2.1	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.2.1	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.2.1	HBase-Befehlszeilen-Client.
hbase-rest-server	1.2.1	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.2.1	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	1.0.0-amzn-6	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	1.0.0-amzn-6	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).

Komponente	Version	Beschreibung
hcatalog-webhcat-server	1.0.0-amzn-6	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	1.0.0-amzn-6	Hive-Befehlszeilen-Client.
hive-metastore-server	1.0.0-amzn-6	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server	1.0.0-amzn-6	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.7.1-amzn-7	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.46	MySQL-Datenbankserver.
oozie-client	4.2.0	Oozie-Befehlszeilen-Client.
oozie-server	4.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.7.0-HBase-1.2	Die Phoenix-Bibliotheken für den Server und den Client

Komponente	Version	Beschreibung
phoenix-query-server	4.7.0-HBase-1.2	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.148	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.148	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.14.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	1.6.2	Spark-Befehlszeilen-Clients.
spark-history-server	1.6.2	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	1.6.2	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	1.6.2	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.3	Die Tez-YARN-Anwendung und -Bibliotheken.



Komponente	Version	Beschreibung
webserver	2.4.23	Apache HTTP-Server.
zeppelin-server	0.5.6-incubating	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.8	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.8	ZooKeeper Befehlszeilenclient.

#### 4.7.4 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

##### emr-4.7.4-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.

Klassifizierungen	Beschreibung
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der .xml-Datei von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.

Klassifizierungen	Beschreibung
hcatalog-webhcat-log4j	Ändert die Werte in der log4j.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j	Ändern Sie die Werte in der Datei 4j.properties von hive-exec-log Hive.
hive-log4j	Ändert die Werte in der hive-log4j.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.

Klassifizierungen	Beschreibung
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.

Klassifizierungen	Beschreibung
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon EMR Version 4.7.3

### 4.7.3 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) und [ZooKeeper-Sandbox](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

#### Informationen zur Anwendungsversion

	emr-4.7.3	emr-4.7.2	emr-4.7.1	emr-4.7.0
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1,10,75
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-

	emr-4.7.3	emr-4.7.2	emr-4.7.1	emr-4.7.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.1	1.2.1	1.2.1	1.2.1
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.2	2.7.2	2.7.2	2.7.2
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.0	0.12.0
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.14.0	0,14,0	0,14,0	0,14,0
Presto	-	-	-	-
Presto-Sandbox	0.148	0.148	0.147	0.147
Spark	1.6.2	1.6.2	1.6.1	1.6.1

	emr-4.7.3	emr-4.7.2	emr-4.7.1	emr-4.7.0
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.3	0.8.3	0.8.3	0.8.3
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.5.6	0.5.6	0.5.6	0.5.6
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.8	3.4.8	3.4.8	3.4.8

## 4.7.3 Versionshinweise

### 4.7.3 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die



Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als 2.2-amzn-2 aufgeführt.

Komponente	Version	Beschreibung
emr-ddb	3.2.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.1.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.2.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.4.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.8.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den

Komponente	Version	Beschreibung
		Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.2-amzn-3	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.2-amzn-3	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.2-amzn-3	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.7.2-amzn-3	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httfs-server	2.7.2-amzn-3	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.7.2-amzn-3	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.7.2-amzn-3	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.2-amzn-3	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.

Komponente	Version	Beschreibung
hadoop-yarn-resourcemanager	2.7.2-amzn-3	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.2-amzn-3	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.2.1	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.2.1	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.2.1	HBase-Befehlszeilen-Client.
hbase-rest-server	1.2.1	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.2.1	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	1.0.0-amzn-6	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	1.0.0-amzn-6	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).

Komponente	Version	Beschreibung
hcatalog-webhcat-server	1.0.0-amzn-6	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	1.0.0-amzn-6	Hive-Befehlszeilen-Client.
hive-metastore-server	1.0.0-amzn-6	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server	1.0.0-amzn-6	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.7.1-amzn-7	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.46	MySQL-Datenbankserver.
oozie-client	4.2.0	Oozie-Befehlszeilen-Client.
oozie-server	4.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.7.0-HBase-1.2	Die Phoenix-Bibliotheken für den Server und den Client

Komponente	Version	Beschreibung
phoenix-query-server	4.7.0-HBase-1.2	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.148	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.148	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.14.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	1.6.2	Spark-Befehlszeilen-Clients.
spark-history-server	1.6.2	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	1.6.2	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	1.6.2	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.3	Die Tez-YARN-Anwendung und -Bibliotheken.

Komponente	Version	Beschreibung
webserver	2.4.23	Apache HTTP-Server.
zeppelin-server	0.5.6-incubating	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.8	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.8	ZooKeeper Befehlszeilenclient.

### 4.7.3 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

#### emr-4.7.3-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.

Klassifizierungen	Beschreibung
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der .xml-Datei von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.

Klassifizierungen	Beschreibung
hcatalog-webhcat-log4j	Ändert die Werte in der log4j.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j	Ändern Sie die Werte in der Datei 4j.properties von hive-exec-log Hive.
hive-log4j	Ändert die Werte in der hive-log4j.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.



Klassifizierungen	Beschreibung
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.

Klassifizierungen	Beschreibung
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 4.7.2

### 4.7.2 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) und [ZooKeeper-Sandbox](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

#### Informationen zur Anwendungsversion

	emr-4.7.2	emr-4.7.1	emr-4.7.0	emr-4.6.1
AWS SDK for Java	1.10.75	1,10,75	1,10,75	1.10,27
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-

	emr-4.7.2	emr-4.7.1	emr-4.7.0	emr-4.6.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.1	1.2.1	1.2.1	1.2.0
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.2	2.7.2	2.7.2	2.7.2
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.0	0.12.0	0.11.1
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	-
Pig	0.14.0	0,14,0	0,14,0	0,14,0
Presto	-	-	-	-
Presto-Sandbox	0.148	0.147	0.147	0.143
Spark	1.6.2	1.6.1	1.6.1	1.6.1

	emr-4.7.2	emr-4.7.1	emr-4.7.0	emr-4.6.1
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.3	0.8.3	0.8.3	-
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.5.6	0.5.6	0.5.6	0.5.6
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.8	3.4.8	3.4.8	3.4.8

## 4.7.2 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon EMR 4.7.2.

Veröffentlichungsdatum: 15. Juli 2016

### Features

- Upgrade auf Mahout 0.12.2
- Upgrade auf Presto 0.148
- Upgrade auf Spark 1.6.2
- Sie können jetzt eine `AWSCredentialsProvider` für die Verwendung mit EMRFS erstellen, indem Sie einen URI als Parameter verwenden. Weitere Informationen finden Sie unter [Erstellen eines `AWSCredentialsProvider` für EMRFS](#).

- EMRFS ermöglicht Benutzern nun die Konfiguration eines benutzerdefinierten DynamoDB -Endpunkts für die Metadaten ihrer konsistenten Ansicht mit der Eigenschaft `fs.s3.consistent.dynamodb.endpoint` in `emrfs-site.xml`.
- Es wurde ein Skript in `/usr/bin` mit dem Namen `spark-example` hinzugefügt, das `/usr/lib/spark/spark/bin/run-example` umschließt, sodass Sie Beispiele direkt ausführen können. Um beispielsweise das SparkPi Beispiel auszuführen, das in der Spark-Distribution enthalten ist, können Sie es `spark-example SparkPi 100` von der Befehlszeile aus ausführen oder es `command-runner.jar` als Schritt in der API verwenden.

Bekannte Probleme aus den früheren Versionen, die behoben wurden

- Es wurde ein Problem behoben, bei dem Oozie die Datei `spark-assembly.jar` nicht am richtigen Speicherort ablegte, wenn Spark auch installiert war. Dies führte dazu, dass Spark-Anwendungen nicht mit Oozie gestartet werden konnten.
- Es wurde ein Problem mit Spark Log4j-basierter Protokollierung in YARN-Containern behoben.

## 4.7.2-Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
emr-ddb	3.2.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.1.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.2.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.4.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.8.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.

Komponente	Version	Beschreibung
hadoop-client	2.7.2-amzn-3	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.2-amzn-3	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.2-amzn-3	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.7.2-amzn-3	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httpfs-server	2.7.2-amzn-3	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.7.2-amzn-3	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.7.2-amzn-3	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.2-amzn-3	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.2-amzn-3	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.



Komponente	Version	Beschreibung
hadoop-yarn-timeline-server	2.7.2-amzn-3	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.2.1	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.2.1	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.2.1	HBase-Befehlszeilen-Client.
hbase-rest-server	1.2.1	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.2.1	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	1.0.0-amzn-6	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	1.0.0-amzn-6	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	1.0.0-amzn-6	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.

Komponente	Version	Beschreibung
hive-client	1.0.0-amzn-6	Hive-Befehlszeilen-Client.
hive-metastore-server	1.0.0-amzn-6	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server	1.0.0-amzn-6	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.7.1-amzn-7	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.2	Bibliothek für Machine Learning.
mysql-server	5.5.46	MySQL-Datenbankserver.
oozie-client	4.2.0	Oozie-Befehlszeilen-Client.
oozie-server	4.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.7.0-HBase-1.2	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.7.0-HBase-1.2	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.

Komponente	Version	Beschreibung
presto-coordinator	0.148	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.148	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.14.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	1.6.2	Spark-Befehlszeilen-Clients.
spark-history-server	1.6.2	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	1.6.2	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	1.6.2	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.3	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.23	Apache HTTP-Server.
zeppelin-server	0.5.6-incubating	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.

Komponente	Version	Beschreibung
zookeeper-server	3.4.8	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.8	ZooKeeper Befehlszeilenclient.

## 4.7.2 Klassifizierungen der Konfiguration

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

### emr-4.7.2-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Hadoop.
hadoop-ssl-server	Ändert die SSL-Server-Konfiguration in Hadoop.

Klassifizierungen	Beschreibung
hadoop-ssl-client	Ändert die SSL-Client-Konfiguration in Hadoop.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der .xml-Datei von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j	Ändert die Werte in der log4j.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.

Klassifizierungen	Beschreibung
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j	Ändern Sie die Werte in der Datei 4j.properties von hive-exec-log Hive.
hive-log4j	Ändert die Werte in der hive-log4j.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.

Klassifizierungen	Beschreibung
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.

Klassifizierungen	Beschreibung
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.



## Amazon-EMR-Version 4.7.1

### 4.7.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) und [ZooKeeper-Sandbox](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

#### Informationen zur Anwendungsversion

	emr-4.7.1	emr-4.7.0	emr-4.6.1	emr-4.6.0
AWS SDK for Java	1.10.75	1,10,75	1.10,27	1.10.27
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-

	emr-4.7.1	emr-4.7.0	emr-4.6.1	emr-4.6.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.1	1.2.1	1.2.0	1.2.0
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.2	2.7.2	2.7.2	2.7.2
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.0	0.12.0	0.11.1	0.11.1
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	-	-
Pig	0.14.0	0,14,0	0,14,0	0,14,0
Presto	-	-	-	-
Presto-Sandbox	0.147	0.147	0.143	0.143
Spark	1.6.1	1.6.1	1.6.1	1.6.1

	emr-4.7.1	emr-4.7.0	emr-4.6.1	emr-4.6.0
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.3	0.8.3	-	-
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.5.6	0.5.6	0.5.6	0.5.6
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.8	3.4.8	3.4.8	3.4.8

## 4.7.1 Versionshinweise

Die folgenden Versionshinweise enthalten Informationen über Amazon EMR 4.7.1.

Veröffentlichungsdatum: 10. Juni 2016

Bekannte Probleme aus den früheren Versionen, die behoben wurden

- Es wurde ein Problem behoben, das die Startup-Zeit von Clustern verlängerte, die in einer VPC mit privaten Subnetzen gestartet wurden. Der Fehler wirkte sich nur auf Cluster aus, die mit Amazon EMR Version 4.7.0 gestartet wurden.
- Es wurde ein Problem behoben, bei dem das Auflisten von Dateien in Amazon EMR für Cluster, die mit Amazon EMR Version 4.7.0 gestartet wurden, nicht ordnungsgemäß durchgeführt wurde.

## 4.7.1 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>emr-ddb</code>	3.1.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.0.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.2.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-s3-dist-cp</code>	2.4.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
<code>emrfs</code>	2.7.1	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>ganglia-monitor</code>	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus

Komponente	Version	Beschreibung
		dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.2-amzn-2	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.2-amzn-2	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.2-amzn-2	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.7.2-amzn-2	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httpfs-server	2.7.2-amzn-2	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.7.2-amzn-2	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider

Komponente	Version	Beschreibung
hadoop-mapred	2.7.2-amzn-2	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.2-amzn-2	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.2-amzn-2	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hadoop-yarn-timeline-server	2.7.2-amzn-2	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.2.1	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.2.1	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.2.1	HBase-Befehlszeilen-Client.
hbase-rest-server	1.2.1	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.2.1	Service, der einen Thrift-Endpunkt für HBase bereitstellt.

Komponente	Version	Beschreibung
hcatalog-client	1.0.0-amzn-5	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	1.0.0-amzn-5	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	1.0.0-amzn-5	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	1.0.0-amzn-5	Hive-Befehlszeilen-Client.
hive-metastore-server	1.0.0-amzn-5	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server	1.0.0-amzn-5	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.7.1-amzn-7	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.0	Bibliothek für Machine Learning.
mysql-server	5.5.46	MySQL-Datenbankserver.
oozie-client	4.2.0	Oozie-Befehlszeilen-Client.

Komponente	Version	Beschreibung
oozie-server	4.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.7.0-HBase-1.2	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.7.0-HBase-1.2	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.
presto-coordinator	0.147	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.147	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.14.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	1.6.1	Spark-Befehlszeilen-Clients.
spark-history-server	1.6.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	1.6.1	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	1.6.1	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.



Komponente	Version	Beschreibung
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.3	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.18	Apache HTTP-Server.
zeppelin-server	0.5.6-incubating	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.8	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.8	ZooKeeper Befehlszeilenclient.

#### 4.7.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

##### emr-4.7.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.

Klassifizierungen	Beschreibung
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der Datei .xml von HCatalog. proto-hive-site

Klassifizierungen	Beschreibung
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j	Ändert die Werte in der log4j.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j	Ändern Sie die Werte in der Datei 4j.properties von hive-exec-log Hive.
hive-log4j	Ändert die Werte in der hive-log4j.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.

Klassifizierungen	Beschreibung
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.

Klassifizierungen	Beschreibung
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 4.7.0

### 4.7.0 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) und [ZooKeeper-Sandbox](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

### Informationen zur Anwendungsversion

	emr-4.7.0	emr-4.6.1	emr-4.6.0	emr-4.5.0
AWS SDK for Java	1.10.75	1.10,27	1.10.27	1.10.27
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt

	emr-4.7.0	emr-4.6.1	emr-4.6.0	emr-4.5.0
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.1	1.2.0	1.2.0	-
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.2	2.7.2	2.7.2	2.7.2
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.0	0.11.1	0.11.1	0.11.1
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	-	-	-

	emr-4.7.0	emr-4.6.1	emr-4.6.0	emr-4.5.0
Pig	0.14.0	0,14,0	0,14,0	0,14,0
Presto	-	-	-	-
Presto-Sandbox	0.147	0.143	0.143	0.140
Spark	1.6.1	1.6.1	1.6.1	1.6.1
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.3	-	-	-
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.5.6	0.5.6	0.5.6	0.5.6
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.8	3.4.8	3.4.8	-

## Versionshinweise zu 4.7.0

### Important

Amazon EMR 4.7.0 ist veraltet. Verwenden Sie stattdessen Amazon EMR 4.7.1 oder höher.

Veröffentlichungsdatum: 2. Juni 2016



## Features

- Apache Phoenix 4.7.0 wurde hinzugefügt.
- Apache Tez 0.8.3 wurde hinzugefügt.
- Upgrade auf Hbase 1.2.1
- Upgrade auf Mahout 0.12.0
- Upgrade auf Presto 0.147
- Auf 1.10.75 AWS SDK for Java aktualisiert
- Das letzte Flag wurde von der Eigenschaft `mapreduce.cluster.local.dir` in `mapred-site.xml` entfernt, damit Benutzer Pig im lokalen Modus ausführen können.
- Amazon-Redshift-JDBC-Treiber verfügbar auf dem Cluster

Amazon Redshift JDBC-Treiber sind jetzt enthalten unter `/usr/share/aws/redshift/jdbc`, `/usr/share/aws/redshift/jdbc/RedshiftJDBC41.jar` ist der JDBC 4.1-kompatible Amazon Redshift-Treiber und `/usr/share/aws/redshift/jdbc/RedshiftJDBC4.jar` ist der JDBC-4.0-kompatible Amazon Redshift-Treiber. Weitere Informationen finden Sie unter [Konfigurieren von JDBC-Verbindungen](#) im Amazon-Redshift-Verwaltungshandbuch.

- Java 8

Außer für Presto ist OpenJDK 1.7 das Standard-JDK, das für alle Anwendungen verwendet wird. Es ist jedoch sowohl OpenJDK 1.7 als auch 1.8 installiert. Informationen zum Festlegen von `JAVA_HOME` für Anwendungen finden Sie unter [Konfigurieren von Anwendungen zur Verwendung von Java 8](#).

## Bekannte Probleme aus den früheren Versionen, die behoben wurden

- Es wurde ein Kernel-Problem behoben, das sich signifikant auf die Leistung für durchsatzoptimierte HDD (st1)-EBS-Volumes für Amazon EMR in emr-4.6.0 auswirkte.
- Es wurde ein Problem behoben, bei dem ein Cluster fehlschlug, wenn eine HDFS-Verschlüsselungszone angegeben wurde, ohne Hadoop als Anwendung auszuwählen.
- Die HDFS-Standardschreibrichtlinie wurde von RoundRobin in `AvailableSpaceVolumeChoosingPolicy` geändert. Einige Volumes wurden mit der RoundRobin Konfiguration nicht richtig genutzt, was zu fehlerhaften Kernknoten und einem unzuverlässigen HDFS führte.

- Es wurde ein Problem mit der EMRFS-Befehlszeilenschnittstelle behoben, das zu einer Ausnahme beim Erstellen der standardmäßigen DynamoDB-Metadatentabelle für konsistente Ansichten führte.
- Es wird ein Deadlock-Problem in EMRFS behoben, das bei mehrteiligen Umbenennungs- und Kopieroperationen auftreten konnte.
- Es wurde ein Problem mit EMRFS behoben, das dazu führte, dass die CopyPart Standardgröße 5 MB betrug. Der Standard ist jetzt ordnungsgemäß auf 128 MB eingestellt.
- Es wurde ein Problem mit der Zeppelin-Startkonfiguration behoben, das möglicherweise das Anhalten des Service verhinderte.
- Es wurde ein Problem mit Spark und Zeppelin behoben, das Sie daran hindert, das `s3a://`-URI-Schema zu verwenden, da `/usr/lib/hadoop/hadoop-aws.jar` nicht ordnungsgemäß in den jeweiligen Klassenpfad geladen wurde.
- [HUE-2484](#) wurde rückportiert.
- Es wurde ein [Commit](#) von Hue 3.9.0 (JIRA nicht vorhanden) rückportiert, um ein Problem mit dem HBase-Browserbeispiel zu beheben.
- [HIVE-9073](#) wurde rückportiert.

## 4.7.0-Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
emr-ddb	3.1.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.0.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.2.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.4.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.7.1	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.

Komponente	Version	Beschreibung
hadoop-client	2.7.2-amzn-2	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.2-amzn-2	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.2-amzn-2	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.7.2-amzn-2	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httpfs-server	2.7.2-amzn-2	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.7.2-amzn-2	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.7.2-amzn-2	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.2-amzn-2	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.2-amzn-2	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.

Komponente	Version	Beschreibung
hadoop-yarn-timeline-server	2.7.2-amzn-2	Service für das Abrufen von aktuellen und historischen Informationen für YARN-Anwendungen.
hbase-hmaster	1.2.1	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.2.1	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.2.1	HBase-Befehlszeilen-Client.
hbase-rest-server	1.2.1	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.2.1	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	1.0.0-amzn-5	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	1.0.0-amzn-5	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	1.0.0-amzn-5	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.

Komponente	Version	Beschreibung
hive-client	1.0.0-amzn-5	Hive-Befehlszeilen-Client.
hive-metastore-server	1.0.0-amzn-5	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server	1.0.0-amzn-5	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.7.1-amzn-7	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.12.0	Bibliothek für Machine Learning.
mysql-server	5.5.46	MySQL-Datenbankserver.
oozie-client	4.2.0	Oozie-Befehlszeilen-Client.
oozie-server	4.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
phoenix-library	4.7.0-HBase-1.2	Die Phoenix-Bibliotheken für den Server und den Client
phoenix-query-server	4.7.0-HBase-1.2	Ein schlanker Server für den Zugriff auf JDBC und Protokollpuffer sowie den Zugriff auf die Avatica-API über das JSON-Format.

Komponente	Version	Beschreibung
presto-coordinator	0.147	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.147	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.14.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	1.6.1	Spark-Befehlszeilen-Clients.
spark-history-server	1.6.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	1.6.1	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	1.6.1	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
tez-on-yarn	0.8.3	Die Tez-YARN-Anwendung und -Bibliotheken.
webserver	2.4.18	Apache HTTP-Server.
zeppelin-server	0.5.6-incubating	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.

Komponente	Version	Beschreibung
zookeeper-server	3.4.8	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.8	ZooKeeper Befehlszeilenclient.

#### 4.7.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

##### emr-4.7.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Hadoop.
hbase-env	Ändert die Werte in der HBase-Umgebung.



Klassifizierungen	Beschreibung
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der .xml-Datei von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j	Ändert die Werte in der log4j.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j	Ändern Sie die Werte in der Datei 4j.properties von hive-exec-log Hive.

Klassifizierungen	Beschreibung
hive-log4j	Ändert die Werte in der hive-log4j.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.

Klassifizierungen	Beschreibung
phoenix-hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in Phoenix.
phoenix-hbase-site	Ändert die Werte in der hbase-site.xml-Datei in Phoenix.
phoenix-log4j	Ändert die Werte in der log4j.properties-Datei in Phoenix.
phoenix-metrics	Ändert die Werte in der hadoop-metrics2-phoenix.properties-Datei in Phoenix.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.

Klassifizierungen	Beschreibung
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
tez-site	Ändert die Werte in der tez-site.xml-Datei in Tez.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon EMR Version 4.6.1

### 4.6.1 Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Zeppelin-Sandbox](#), und [ZooKeeper-Sandbox](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

#### Informationen zur Anwendungsversion

	emr-4.6.1	emr-4.6.0	emr-4.5.0	emr-4.4.0
AWS SDK for Java	1.10.27	1.10.27	1.10.27	1.10.27
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.0	1.2.0	-	-
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.2	2.7.2	2.7.2	2.7.1

	emr-4.6.1	emr-4.6.0	emr-4.5.0	emr-4.4.0
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.11.1	0.11.1	0.11.1	0.11.1
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	-	-	-	-
Pig	0.14.0	0,14,0	0,14,0	0,14,0
Presto	-	-	-	-
Presto-Sandbox	0.143	0.143	0.140	0.136
Spark	1.6.1	1.6.1	1.6.1	1.6.0
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	-	-	-	-

	emr-4.6.1	emr-4.6.0	emr-4.5.0	emr-4.4.0
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.5.6	0.5.6	0.5.6	0.5.6
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.8	3.4.8	-	-

## 4.6.1 Versionshinweise

### 4.6.1 Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
emr-ddb	3.0.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.

Komponente	Version	Beschreibung
emr-goodies	2.0.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.1.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.3.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.6.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.2-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".



Komponente	Version	Beschreibung
hadoop-hdfs-datanode	2.7.2-amzn-1	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.2-amzn-1	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.7.2-amzn-1	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httpfs-server	2.7.2-amzn-1	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.7.2-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.7.2-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.2-amzn-1	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resource-manager	2.7.2-amzn-1	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.

Komponente	Version	Beschreibung
hbase-hmaster	1.2.0	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.2.0	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.2.0	HBase-Befehlszeilen-Client.
hbase-rest-server	1.2.0	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.
hbase-thrift-server	1.2.0	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	1.0.0-amzn-4	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	1.0.0-amzn-4	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	1.0.0-amzn-4	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	1.0.0-amzn-4	Hive-Befehlszeilen-Client.

Komponente	Version	Beschreibung
hive-metastore-server	1.0.0-amzn-4	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server	1.0.0-amzn-4	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.7.1-amzn-6	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.11.1	Bibliothek für Machine Learning.
mysql-server	5.5	MySQL-Datenbankserver.
oozie-client	4.2.0	Oozie-Befehlszeilen-Client.
oozie-server	4.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
presto-coordinator	0.143	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.143	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.14.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	1.6.1	Spark-Befehlszeilen-Clients.

Komponente	Version	Beschreibung
spark-history-server	1.6.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	1.6.1	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	1.6.1	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
webserver	2.4	Apache HTTP-Server.
zeppelin-server	0.5.6-incubating	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.8	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.8	ZooKeeper Befehlszeilenclient.

## 4.6.1 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

### emr-4.6.1-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Hadoop.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der <code>hbase-log4j.properties</code> -Datei in HBase.
hbase-metrics	Ändert die Werte in der <code>hadoop-metrics2-hbase.properties</code> -Datei in HBase.
hbase-policy	Ändert die Werte in der <code>hbase-policy.xml</code> -Datei in HBase.
hbase-site	Ändert die Werte in der <code>hbase-site.xml</code> -Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.

Klassifizierungen	Beschreibung
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der XML-Datei von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j	Ändert die Werte in der log4j.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j	Ändern Sie die Werte in der Datei 4j.properties von hive-exec-log Hive.
hive-log4j	Ändert die Werte in der hive-log4j.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.

Klassifizierungen	Beschreibung
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.



## Amazon-EMR-Version 4.6.0

### 4.6.0-Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Zeppelin-Sandbox](#), und [ZooKeeper-Sandbox](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

#### Informationen zur Anwendungsversion

	emr-4.6.0	emr-4.5.0	emr-4.4.0	emr-4.3.0
AWS SDK for Java	1.10.27	1.10.27	1.10.27	1.10.27
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-

	emr-4.6.0	emr-4.5.0	emr-4.4.0	emr-4.3.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.0	-	-	-
HCatalog	1.0.0	1.0.0	1.0.0	-
Hadoop	2.7.2	2.7.2	2.7.1	2.7.1
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.11.1	0.11.1	0.11.1	0.11.0
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	-	-	-	-
Pig	0.14.0	0,14,0	0,14,0	0,14,0
Presto	-	-	-	-
Presto-Sandbox	0.143	0.140	0.136	0.130
Spark	1.6.1	1.6.1	1.6.0	1.6.0

	emr-4.6.0	emr-4.5.0	emr-4.4.0	emr-4.3.0
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	-
TensorFlow	-	-	-	-
Tez	-	-	-	-
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.5.6	0.5.6	0.5.6	0.5.5
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.8	-	-	-

## Versionshinweise zu 4.6.0

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 4.6.0.

- HBase 1.2.0 wurde hinzugefügt.
- ZooKeeper-Sandbox 3.4.8 wurde hinzugefügt.
- Upgrade auf Presto-Sandbox 0.143
- Amazon-EMR-Versionen basieren jetzt auf Amazon Linux 2016.03.0. Weitere Informationen finden Sie unter <https://aws.amazon.com/amazon-linux-ami/2016.03-release-notes/>.
- Problem mit den durchsatzoptimierten HDD (st1)-EBS-Volumen-Typen

Ein Problem im Linux-Kernel Versionen 4.2 und höher wirkte sich signifikant auf die Leistung der durchsatzoptimierten HDD (st1)-EBS-Volumen für EMR aus. Diese Version (emr-4.6.0) verwendet die Kernel-Version 4.4.5 und ist daher betroffen. Es wird empfohlen, emr-4.6.0 nicht zu verwenden, wenn Sie st1-EBS-Volumen einsetzen möchten. Sie können Amazon-EMR-Versionen emr-4.5.0

oder früher mit `st1` ohne Auswirkungen verwenden. Außerdem stellen wir die Problembeseitigung mit späteren Versionen zur Verfügung.

- Python-Standard-Einstellungen

Python 3.4 ist jetzt standardmäßig installiert, Python 2.7 bleibt jedoch der Systemstandard. Sie können Python 3.4 mit einer Bootstrap-Aktion als Systemstandard konfigurieren. Sie können die Konfigurations-API verwenden, um den `PYSPARK_PYTHON`-Export `/usr/bin/python3.4` in der `spark-env` Klassifizierung auf festzulegen, um die Python-Version zu beeinflussen, die verwendet wird. PySpark

- Java 8

Außer für Presto ist OpenJDK 1.7 das Standard-JDK, das für alle Anwendungen verwendet wird. Es ist jedoch sowohl OpenJDK 1.7 als auch 1.8 installiert. Informationen zum Festlegen von `JAVA_HOME` für Anwendungen finden Sie unter [Konfigurieren von Anwendungen zur Verwendung von Java 8](#).

Bekannte Probleme aus den früheren Versionen, die behoben wurden

- Es wurde ein Problem behoben, bei dem die Anwendungsbereitstellung manchmal wegen eines generierten Passwort nach dem Zufallsprinzip ausfiel.
- Bisher war `mysql` auf allen Knoten installiert. Jetzt ist es nur auf der Master-Instance installiert und nur dann, wenn die gewählte Anwendung `mysql-server` als Komponente enthält. Derzeit enthalten die folgenden Anwendungen die Komponente `mysql-server`: HCatalog, Hive, Hue, Presto-Sandbox und Sqoop-Sandbox.
- Der Wert `yarn.scheduler.maximum-allocation-vcores` wurde von der Standardeinstellung 32 auf 80 geändert. Dadurch wurde in `emr-4.4.0` eingeführtes Problem behoben, das hauptsächlich mit Spark auftritt bei der Verwendung der Option `maximizeResourceAllocation` in einem Cluster, dessen Core-Instance-Typ zu den wenigen großen Instance-Typen gehört, deren YARN `vcores` höher als 32 festgelegt sind. Und zwar waren `c4.8xlarge`, `cc2.8xlarge`, `hs1.8xlarge`, `i2.8xlarge`, `m2.4xlarge`, `r3.8xlarge`, `d2.8xlarge` oder `m4.10xlarge` von dem Problem betroffen.
- `s3-dist-cp` verwendet jetzt für alle Amazon-S3-Nominierungen EMRFS und keine Stufen eines temporären HDFS-Verzeichnisses mehr.
- Es wurde ein Problem mit der Ausnahmebehandlung für die clientseitigen Verschlüsselung mehrteiliger Uploads behoben.
- Es wurde eine Option hinzugefügt, mit der Benutzer die Amazon-S3-Speicherklasse ändern können. Standardmäßig ist diese Einstellung `STANDARD`. Die

Konfigurationsklassifizierungseinstellung `emrfs-site` ist `fs.s3.storageClass` und die möglichen Werte sind `STANDARD`, `STANDARD_IA` und `REDUCED_REDUNDANCY`. Weitere Informationen zu Speicherklassen finden Sie unter [Speicherklassen](#) im Benutzerhandbuch zu Amazon Simple Storage Service.

#### 4.6.0-Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>emr-ddb</code>	3.0.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.0.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.1.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-s3-dist-cp</code>	2.3.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.

Komponente	Version	Beschreibung
emrfs	2.6.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.2-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.2-amzn-1	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.2-amzn-1	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.7.2-amzn-1	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.

Komponente	Version	Beschreibung
hadoop-httfs-server	2.7.2-amzn-1	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.7.2-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.7.2-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.2-amzn-1	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.2-amzn-1	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hbase-hmaster	1.2.0	Service für einen HBase-Cluster, der für die Koordination der Regionen und die Ausführung von administrativen Befehlen zuständig ist.
hbase-region-server	1.2.0	Service für die Bereitstellung einer oder mehrerer HBase-Regionen.
hbase-client	1.2.0	HBase-Befehlszeilen-Client.
hbase-rest-server	1.2.0	Service, der einen RESTful-HTTP-Endpunkt für HBase bereitstellt.

Komponente	Version	Beschreibung
hbase-thrift-server	1.2.0	Service, der einen Thrift-Endpunkt für HBase bereitstellt.
hcatalog-client	1.0.0-amzn-4	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	1.0.0-amzn-4	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	1.0.0-amzn-4	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	1.0.0-amzn-4	Hive-Befehlszeilen-Client.
hive-metastore-server	1.0.0-amzn-4	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server	1.0.0-amzn-4	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.7.1-amzn-6	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.11.1	Bibliothek für Machine Learning.
mysql-server	5.5	MySQL-Datenbankserver.



Komponente	Version	Beschreibung
oozie-client	4.2.0	Oozie-Befehlszeilen-Client.
oozie-server	4.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
presto-coordinator	0.143	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.143	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.14.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	1.6.1	Spark-Befehlszeilen-Clients.
spark-history-server	1.6.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	1.6.1	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	1.6.1	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
webserver	2.4	Apache HTTP-Server.

Komponente	Version	Beschreibung
zeppelin-server	0.5.6-incubating	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.
zookeeper-server	3.4.8	Zentraler Service für die Verwaltung von Konfigurationsinformationen, die Benennung, die Bereitstellung verteilter Synchronisierung und die Bereitstellung von Gruppenservices.
zookeeper-client	3.4.8	ZooKeeper Befehlszeilenclient.

#### 4.6.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

##### emr-4.6.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.

Klassifizierungen	Beschreibung
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hbase-env	Ändert die Werte in der HBase-Umgebung.
hbase-log4j	Ändert die Werte in der hbase-log4j.properties-Datei in HBase.
hbase-metrics	Ändert die Werte in der hadoop-metrics2-hbase.properties-Datei in HBase.
hbase-policy	Ändert die Werte in der hbase-policy.xml-Datei in HBase.
hbase-site	Ändert die Werte in der hbase-site.xml-Datei in HBase.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der jndi.properties-Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der .xml-Datei von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j	Ändert die Werte in der log4j.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.

Klassifizierungen	Beschreibung
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j	Ändern Sie die Werte in der Datei 4j.properties von hive-exec-log Hive.
hive-log4j	Ändert die Werte in der hive-log4j.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.

Klassifizierungen	Beschreibung
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.

Klassifizierungen	Beschreibung
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.
zookeeper-config	Ändern Sie die Werte in ZooKeeper der Datei zoo.cfg.
zookeeper-log4j	Ändern Sie die Werte in ZooKeeper der Datei log4j.properties.

## Amazon-EMR-Version 4.5.0

### 4.5.0-Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Ganglia](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#) und [Zeppelin-Sandbox](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

## Informationen zur Anwendungsversion

	emr-4.5.0	emr-4.4.0	emr-4.3.0	emr-4.2.0
AWS SDK for Java	1.10.27	1.10.27	1.10.27	1.10.27
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.6.0
HBase	-	-	-	-
HCatalog	1.0.0	1.0.0	-	-
Hadoop	2.7.2	2.7.1	2.7.1	2.6.0
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-

	emr-4.5.0	emr-4.4.0	emr-4.3.0	emr-4.2.0
MXNet	-	-	-	-
Mahout	0.11.1	0.11.1	0.11.0	0.11.0
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	-	-	-	-
Pig	0.14.0	0,14,0	0,14,0	0,14,0
Presto	-	-	-	-
Presto-Sandbox	0.140	0.136	0.130	0.125
Spark	1.6.1	1.6.0	1.6.0	1.5.2
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	-	-
TensorFlow	-	-	-	-
Tez	-	-	-	-
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.5.6	0.5.6	0.5.5	0,5,5
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	-	-	-	-



## Versionshinweise zu 4.5.0

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 4.5.0.

Veröffentlichungsdatum: 4. April 2016

### Features

- Upgrade auf Spark 1.6.1
- Upgrade auf Hadoop 2.7.2
- Upgrade auf Presto 0.140
- AWS KMS Unterstützung für serverseitige Amazon S3 S3-Verschlüsselung hinzugefügt.

Bekannte Probleme aus den früheren Versionen, die behoben wurden

- Es wurde ein Problem behoben, bei dem MySQL- und Apache-Server nicht gestartet wurden, nachdem ein Knoten neu gestartet wurde.
- Es wurde ein Problem behoben, bei dem IMPORT nicht ordnungsgemäß mit nicht partitionierte Tabellen in Amazon S3 funktionierte.
- Es wurde ein Problem mit Presto behoben, bei dem das Staging-Verzeichnis beim Schreiben in Hive-Tabellen `/mnt/tmp` anstatt `/tmp` lauten muss.

## 4.5.0-Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
emr-ddb	3.0.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.0.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.1.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.2.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.5.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.

Komponente	Version	Beschreibung
hadoop-client	2.7.2-amzn-0	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.2-amzn-0	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.2-amzn-0	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.7.2-amzn-0	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httfs-server	2.7.2-amzn-0	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.7.2-amzn-0	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.7.2-amzn-0	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.2-amzn-0	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.7.2-amzn-0	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.

Komponente	Version	Beschreibung
hcatalog-client	1.0.0-amzn-4	Der "hcat"-Befehlszeilen-Client für das Bearbeiten des hcatalog-Servers.
hcatalog-server	1.0.0-amzn-4	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	1.0.0-amzn-4	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	1.0.0-amzn-4	Hive-Befehlszeilen-Client.
hive-metastore-server	1.0.0-amzn-4	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server	1.0.0-amzn-4	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.7.1-amzn-5	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.11.1	Bibliothek für Machine Learning.
mysql-server	5.5	MySQL-Datenbankserver.
oozie-client	4.2.0	Oozie-Befehlszeilen-Client.

Komponente	Version	Beschreibung
oozie-server	4.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
presto-coordinator	0.140	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.140	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.14.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	1.6.1	Spark-Befehlszeilen-Clients.
spark-history-server	1.6.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	1.6.1	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	1.6.1	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
webserver	2.4	Apache HTTP-Server.
zeppelin-server	0.5.6-incubating	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.

## 4.5.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

### emr-4.5.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Hadoop.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der <code>hdfs-site.xml</code> -Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der <code>jndi.properties</code> -Datei von HCatalog.
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der <code>.xml</code> -Datei von HCatalog. <code>proto-hive-site</code>
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.

Klassifizierungen	Beschreibung
hcatalog-webhcat-log4j	Ändert die Werte in der log4j.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j	Ändern Sie die Werte in der Datei 4j.properties von hive-exec-log Hive.
hive-log4j	Ändert die Werte in der hive-log4j.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.

Klassifizierungen	Beschreibung
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.



Klassifizierungen	Beschreibung
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oraoop-site	Ändern Sie die Werte in der Datei oraoop-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.

## Amazon-EMR-Version 4.4.0

### 4.4.0-Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Ganglia](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#) und [Zeppelin-Sandbox](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

## Informationen zur Anwendungsversion

	emr-4.4.0	emr-4.3.0	emr-4.2.0	emr-4.1.0
AWS SDK for Java	1.10.27	1.10.27	1.10.27	Nicht nachverfolgt
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.6.0	-
HBase	-	-	-	-
HCatalog	1.0.0	-	-	-
Hadoop	2.7.1	2.7.1	2.6.0	2.6.0
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-

	emr-4.4.0	emr-4.3.0	emr-4.2.0	emr-4.1.0
MXNet	-	-	-	-
Mahout	0.11.1	0.11.0	0.11.0	0.11.0
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.0.1
Phoenix	-	-	-	-
Pig	0,14,0	0,14,0	0,14,0	0,14,0
Presto	-	-	-	-
Presto-Sandbox	0.136	0.130	0.125	0.119
Spark	1.6.0	1.6.0	1.5.2	1.5.0
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	-	-	-
TensorFlow	-	-	-	-
Tez	-	-	-	-
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.5.6	0.5.5	0,5,5	0.6.0-SNA PSHOT
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	-	-	-	-

## Versionshinweise zu 4.4.0

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 4.4.0.

Veröffentlichungsdatum: 14. März 2016

### Features

- HCatalog 1.0.0 wurde hinzugefügt.
- Sqoop-Sandbox 1.4.6 wurde hinzugefügt.
- Upgrade auf Presto 0.136
- Upgrade auf Zeppelin 0.5.6
- Upgrade auf Mahout 0.11.1
- `dynamicResourceAllocation` wurde standardmäßig aktiviert.
- Es wurde eine Tabelle mit allen Konfigurationsklassifizierungen für die Version hinzugefügt. Weitere Informationen finden Sie in der Konfigurationsklassifizierungstabelle unter [Konfigurieren von Anwendungen](#).

### Bekannte Probleme aus den früheren Versionen, die behoben wurden

- Es wurde ein Problem behoben, bei dem die `maximizeResourceAllocation` Einstellung nicht genug Speicher für ApplicationMaster YARN-Daemons reservierte.
- Es wurde ein Problem mit einer benutzerdefinierten DNS behoben. Wenn Einträge in `resolve.conf` den angegebenen, benutzerdefinierten Einträge vorangestellt werden, können die benutzerdefinierten Einträge nicht aufgelöst werden. Dieses Verhalten wirkte sich auf Cluster in einer VPC aus, wo der Standard-VPC-Nameserver als oberster Eintrag in `resolve.conf` eingefügt wird.
- Es wurde ein Problem behoben, bei dem die Python-Standardversion auf Version 2.7 geändert wurde und boto für diese Version nicht installiert war.
- Es wurde ein Problem behoben, bei dem YARN-Container und Spark-Anwendungen eine eindeutige Ganglia Round Robin-Datenbankdatei (rrd) generierte, wodurch der erste, der Instance angefügte Datenträger vollständig belegt wurde. Aufgrund dieser Fehlerbehebung wurden YARN-Metriken auf Containerebene und Spark-Metriken auf Anwendungsebene deaktiviert.
- Es wurde ein Problem im Protokoll-Pusher behoben, bei dem alle leeren Protokollordner gelöscht wurden. Das Ergebnis war, dass die Hive-CLI keine Protokolle erstellen konnte, da der Protokoll-Pusher den leeren `user`-Ordner unter `/var/log/hive` entfernte.

- Es wurde ein Problem mit Hive-Importen behoben, da sich auf die Partitionierung auswirkte und beim Import zu einem Fehler führte.
- Es wurde ein Problem behoben, bei dem EMRFS und s3-dist-cp Bucket-Namen, die Punkte enthielten, nicht ordnungsgemäß verarbeiteten.
- Es wurde ein Verhalten in EMRFS geändert, damit in Versioning-fähigen Buckets die Markierungsdatei `_$folder$` nicht kontinuierlich erstellt wird. Dies kann zu einer verbesserten Leistung für Versioning-fähige Buckets beitragen.
- Es wurde das Verhalten in EMRFS geändert, sodass Anweisungsdateien nur in Fällen verwendet werden, in denen die clientseitige Verschlüsselung aktiviert ist. Wenn Sie Anweisungsdateien bei Verwendung der clientseitigen Verschlüsselung löschen möchten, können Sie die `emrfs-site.xml`-Eigenschaft, `fs.s3.cse.cryptoStorageMode.deleteInstructionFiles.enabled`, auf "true" festlegen.
- Die YARN-Protokollaggregation wurde geändert, sodass Protokolle am Aggregationsziel zwei Tage lang verwahrt werden. Das Standard-Ziel ist der HDFS-Speicher Ihres Clusters. Wenn Sie diese Dauer ändern möchten, ändern Sie den Wert `yarn.log-aggregation.retain-seconds` mit der Konfigurationsklassifizierung `yarn-site` beim Erstellen Ihres Clusters. Wie immer können Sie Ihre Anwendungsprotokolle beim Erstellen Ihres Clusters mit dem Parameter `log-uri` in Amazon S3 speichern.

## Angewendete Patches

- [HIVE-9655](#)
- [HIVE-9183](#)
- [HADOOP-12810](#)

## 4.4.0-Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-*

amzn-*EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen myapp-component der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als 2.2-amzn-2 aufgeführt.

Komponente	Version	Beschreibung
emr-ddb	3.0.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.0.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.1.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.2.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.4.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagenten.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den

Komponente	Version	Beschreibung
		Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.1-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.1-amzn-1	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.1-amzn-1	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.7.1-amzn-1	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httfs-server	2.7.1-amzn-1	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.7.1-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.7.1-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.1-amzn-1	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.

Komponente	Version	Beschreibung
hadoop-yarn-resourcemanager	2.7.1-amzn-1	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hcatalog-client	1.0.0-amzn-3	Der "hcat"-Befehlszeilen-Client-für das Bearbeiten des hcatalog-Servers.
hcatalog-server	1.0.0-amzn-3	Service, der HCatalog bereitstellt (ein Tabellen- und Speicherverwaltungs-Layer für verteilte Anwendungen).
hcatalog-webhcat-server	1.0.0-amzn-3	HTTP-Endpunkt, der eine REST-Schnittstelle für HCatalog bereitstellt.
hive-client	1.0.0-amzn-3	Hive-Befehlszeilen-Client.
hive-metastore-server	1.0.0-amzn-3	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server	1.0.0-amzn-3	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.7.1-amzn-5	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.11.1	Bibliothek für Machine Learning.



Komponente	Version	Beschreibung
mysql-server	5.5	MySQL-Datenbankserver.
oozie-client	4.2.0	Oozie-Befehlszeilen-Client.
oozie-server	4.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
presto-coordinator	0.136	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.136	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.14.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	1.6.0	Spark-Befehlszeilen-Clients.
spark-history-server	1.6.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	1.6.0	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	1.6.0	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
sqoop-client	1.4.6	Apache Sqoop-Befehlszeilen-Client.
webserver	2.4	Apache HTTP-Server.

Komponente	Version	Beschreibung
zeppelin-server	0.5.6-incubating	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.

#### 4.4.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

##### emr-4.4.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Hadoop.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der <code>hdfs-site.xml</code> -Datei in HDFS.
hcatalog-env	Ändert die Werte in der HCatalog-Umgebung.
hcatalog-server-jndi	Ändert die Werte in der <code>jndi.properties</code> -Datei von HCatalog.

Klassifizierungen	Beschreibung
hcatalog-server-proto-hive-site	Ändern Sie die Werte in der .xml-Datei von HCatalog. proto-hive-site
hcatalog-webhcat-env	Ändert die Werte in der WebHCat-Umgebung von HCatalog.
hcatalog-webhcat-log4j	Ändert die Werte in der log4j.properties-Datei von WebHCat in HCatalog.
hcatalog-webhcat-site	Ändert die Werte in der webhcat-site.xml-Datei von WebHCat in HCatalog.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j	Ändern Sie die Werte in der Datei 4j.properties von hive-exec-log Hive.
hive-log4j	Ändert die Werte in der hive-log4j.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.

Klassifizierungen	Beschreibung
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.

Klassifizierungen	Beschreibung
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
sqoop-env	Ändert die Werte in der Sqoop-Umgebung.
sqoop-oracore-site	Ändern Sie die Werte in der Datei oracore-site.xml OraOop von Sqoop.
sqoop-site	Ändert die Werte in der sqoop-site.xml in Sqoop.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.

## Amazon-EMR-Version 4.3.0

### 4.3.0-Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Ganglia](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Pig](#), [Presto-Sandbox](#), [Spark](#) und [Zeppelin-Sandbox](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)

- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

### Informationen zur Anwendungsversion

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
AWS SDK for Java	1.10.27	1.10.27	Nicht nachverfolgt	Nicht nachverfolgt
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.6.0	-	-
HBase	-	-	-	-
HCatalog	-	-	-	-
Hadoop	2.7.1	2.6.0	2.6.0	2.6.0
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	-
Iceberg	-	-	-	-

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.11.0	0.11.0	0.11.0	0.10.0
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.0.1	-
Phoenix	-	-	-	-
Pig	0,14,0	0,14,0	0,14,0	0,14,0
Presto	-	-	-	-
Presto-Sandbox	0.130	0.125	0.119	-
Spark	1.6.0	1.5.2	1.5.0	1.4.1
Sqoop	-	-	-	-
Sqoop-Sandbox	-	-	-	-
TensorFlow	-	-	-	-
Tez	-	-	-	-
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.5.5	0,5,5	0.6.0-SNA PSHOT	-

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	-	-	-	-

## Versionshinweise zu 4.3.0

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 4.3.0.

Veröffentlichungsdatum: 19. Januar 2016

### Features

- Upgrade auf Hadoop 2.7.1
- Upgrade auf Spark 1.6.0
- Upgrade auf Ganglia 3.7.2
- Upgrade auf Presto 0.130
- Amazon EMR hat einige Änderungen an der Einstellung `spark.dynamicAllocation.enabled` vorgenommen, wenn sie auf „true“ gesetzt wurde. Standardmäßig lautet der Wert „false“. Wenn die Einstellung auf "true" festgelegt ist, wirkt sich dies auf von der Einstellung `maximizeResourceAllocation` vorgenommene Standardeinstellungen aus:
  - Wenn `spark.dynamicAllocation.enabled` auf "true" eingestellt ist, wird `spark.executor.instances` nicht von `maximizeResourceAllocation` festgelegt.
  - Die Einstellung `spark.driver.memory` wird nun basierend auf den Instance-Typen im Cluster auf ähnliche Weise konfiguriert, wie `spark.executors.memory` festgelegt wird. Da die Spark-Treiberanwendung jedoch entweder in der Master-Instance oder in einer der Core-Instances ausgeführt werden kann (z. B. im YARN-Client- bzw. -Cluster-Modus), wird die Einstellung `spark.driver.memory` auf der Grundlage des kleineren Instance-Typs in diesen beiden Instance-Gruppen festgelegt.
  - Die Einstellung `spark.default.parallelism` wird jetzt auf die doppelte Anzahl von CPU-Kernen für YARN Container festgelegt. In früheren Versionen war dies die Hälfte des Werts.
  - Die Berechnungen für den Arbeitsspeicher-Overhead, der für Spark YARN-Prozesse reserviert wurde, wurde präziser angepasst. Dies führt zu einer kleinen Erhöhung des gesamten, verfügbaren Speichers für Spark (d. h. `spark.executor.memory`).



## Bekannte Probleme aus den früheren Versionen, die behoben wurden

- Die YARN-Protokollaggregation ist jetzt standardmäßig aktiviert.
- Es wurde ein Problem behoben, bei dem Protokolle nicht per Push auf den Amazon-S3-Protokoll-Bucket für das Cluster übertragen wurden, wenn die YARN-Protokollaggregation aktiviert war.
- Die Größe von YARN-Containern haben jetzt ein neues Minimum von 32 für alle Knotentypen.
- Es wurde ein Problem mit Ganglia behoben, dass zu übermäßigen Festplatten-E/A-Vorgängen auf dem Primärknoten in großen Clustern führte.
- Es wurde ein Problem behoben, das verhinderte, dass Anwendungsprotokolle per Push-Verfahren an Amazon S3 übertragen wurden, wenn ein Cluster heruntergefahren wird.
- Es wurde ein Problem in der EMRFS-CLI behoben, das dazu führte, dass einige Befehle fehlschlugen.
- Es wurde ein Problem mit Zeppelin behoben, das verhinderte, dass Abhängigkeiten in die Basisdatei geladen wurden. SparkContext
- Es wurde ein Problem behoben, das aus einer Größenanpassung resultierte bei dem Versuch, Instances hinzuzufügen.
- Es wurde ein Problem in Hive behoben, bei dem CREATE TABLE AS SELECT übermäßige Listenaufrufe an Amazon S3 erstellt.
- Es wurde ein Problem behoben, bei dem große Cluster nicht ordnungsgemäß bereitgestellt wurden wenn Hue, Oozie und Ganglia installiert sind.
- Es wurde ein Problem in s3-dist-cp behoben, bei dem ein Beendigungscode gleich Null zurückgegeben wurde, wenn ein Fehler auftrat.

### 4.3.0-Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion-amzn-EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-

Community-Komponente mit dem Namen myapp-component der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als 2.2-amzn-2 aufgeführt.

Komponente	Version	Beschreibung
emr-ddb	3.0.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-goodies	2.0.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
emr-kinesis	3.1.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
emr-s3-dist-cp	2.1.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
emrfs	2.3.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
ganglia-monitor	3.7.2	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.7.2	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.7.1	Webanwendung zum Anzeigen von durch den

Komponente	Version	Beschreibung
		Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.7.1-amzn-0	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.7.1-amzn-0	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.7.1-amzn-0	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.7.1-amzn-0	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httfs-server	2.7.1-amzn-0	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.7.1-amzn-0	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.7.1-amzn-0	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.7.1-amzn-0	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.

Komponente	Version	Beschreibung
hadoop-yarn-resourcemanager	2.7.1-amzn-0	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hive-client	1.0.0-amzn-2	Hive-Befehlszeilen-Client.
hive-metastore-server	1.0.0-amzn-2	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server	1.0.0-amzn-2	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.7.1-amzn-5	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.11.0	Bibliothek für Machine Learning.
mysql-server	5.5	MySQL-Datenbankserver.
oozie-client	4.2.0	Oozie-Befehlszeilen-Client.
oozie-server	4.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
presto-coordinator	0.130	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.

Komponente	Version	Beschreibung
presto-worker	0.130	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.14.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	1.6.0	Spark-Befehlszeilen-Clients.
spark-history-server	1.6.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	1.6.0	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	1.6.0	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
webserver	2.4	Apache HTTP-Server.
zeppelin-server	0.5.5-incubating-amzn-1	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.

### 4.3.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

## emr-4.3.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der capacity-scheduler.xml-Datei in Hadoop.
core-site	Ändert die Werte in der core-site.xml-Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j	Ändern Sie die Werte in der Datei 4j.properties von hive-exec-log Hive.
hive-log4j	Ändert die Werte in der hive-log4j.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.

Klassifizierungen	Beschreibung
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.

Klassifizierungen	Beschreibung
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.

## Amazon-EMR-Version 4.2.0

### 4.2.0-Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Ganglia](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Pig](#), [Presto-Sandbox](#), [Spark](#) und [Zeppelin-Sandbox](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:



- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

### Informationen zur Anwendungsversion

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
AWS SDK for Java	1.10.27	1.10.27	Nicht nachverfolgt	Nicht nachverfolgt
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
Scala	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.6.0	-	-
HBase	-	-	-	-
HCatalog	-	-	-	-
Hadoop	2.7.1	2.6.0	2.6.0	2.6.0
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	-
Iceberg	-	-	-	-

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.11.0	0.11.0	0.11.0	0.10.0
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.0.1	-
Phoenix	-	-	-	-
Pig	0,14,0	0,14,0	0,14,0	0,14,0
Presto	-	-	-	-
Presto-Sandbox	0.130	0.125	0.119	-
Spark	1.6.0	1.5.2	1.5.0	1.4.1
Sqoop	-	-	-	-
Sqoop-Sandbox	-	-	-	-
TensorFlow	-	-	-	-
Tez	-	-	-	-
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.5.5	0,5,5	0.6.0-SNA PSHOT	-

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	-	-	-	-

## Versionshinweise zu 4.2.0

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 4.2.0.

Veröffentlichungsdatum: 18. November 2015

### Features

- Ganglia-Support wurde hinzugefügt.
- Upgrade auf Spark 1.5.2
- Upgrade auf Presto 0.125
- Upgrade auf Oozie 4.2.0
- Upgrade auf Zeppelin 0.5.5
- Auf 1.10.27 AWS SDK for Java aktualisiert

### Bekannte Probleme aus den früheren Versionen, die behoben wurden

- Es wurde ein Problem mit der EMRFS-Befehlszeilenschnittstelle behoben, bei dem der standardmäßige Metadatentabellenname nicht verwendet wurde.
- Es wurde ein Problem behoben, das bei der Verwendung von ORC-gestützten Tabellen in Amazon S3 auftrat.
- Es wurde ein Problem behoben, das bei einem Python-Versionskonflikt in der Spark-Konfiguration auftrat.
- Es wurde ein Problem behoben, bei dem ein YARN-Knotenstatus aufgrund von DNS-Problemen für Cluster in einer VPC keine Daten sendete.
- Es wurde ein Problem behoben, die auftraten, wenn YARN Knoten stilllegte. Das führt dazu, dass Anwendungen hängen blieben oder keine neuen Anwendungen geplant werden konnten.
- Es wurde ein Problem behoben, das auftrat, wenn Cluster mit dem Status TIMED\_OUT\_STARTING beendet wurden.

- Es wurde ein Problem behoben, das auftrat, wenn die EMRFS Scala-Abhängigkeit in andere Builds einbezogen wurde. Die Scala-Abhängigkeit wurde entfernt.

## 4.2.0-Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form `CommunityVersion-amzn-EmrVersion`. Der `EmrVersion` beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>emr-ddb</code>	3.0.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.0.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.1.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-s3-dist-cp</code>	2.0.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.

Komponente	Version	Beschreibung
emrfs	2.2.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
ganglia-monitor	3.6.0	Eingebetteter Ganglia-Agent für Anwendungen aus dem Hadoop-Ökosystem zusammen mit dem Ganglia-Überwachungsagent.
ganglia-metadata-collector	3.6.0	Ganglia-Metadaten-Kollektor zum Aggregieren von Metriken aus Ganglia-Überwachungsagenten.
ganglia-web	3.5.10	Webanwendung zum Anzeigen von durch den Ganglia-Metadaten-Kollektor gesammelten Metriken.
hadoop-client	2.6.0-amzn-2	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
hadoop-hdfs-datanode	2.6.0-amzn-2	HDFS-Service auf Knotenebene zum Speichern von Blöcken.
hadoop-hdfs-library	2.6.0-amzn-2	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.6.0-amzn-2	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.

Komponente	Version	Beschreibung
hadoop-httfs-server	2.6.0-amzn-2	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.6.0-amzn-2	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.6.0-amzn-2	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.6.0-amzn-2	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.6.0-amzn-2	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hive-client	1.0.0-amzn-1	Hive-Befehlszeilen-Client.
hive-metastore-server	1.0.0-amzn-1	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server	1.0.0-amzn-1	Service zur Annahme von Hive-Abfragen als Webanfragen.

Komponente	Version	Beschreibung
hue-server	3.7.1-amzn-5	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.11.0	Bibliothek für Machine Learning.
mysql-server	5.5	MySQL-Datenbankserver.
oozie-client	4.2.0	Oozie-Befehlszeilen-Client.
oozie-server	4.2.0	Service für die Annahme von Oozie Workflow-Anforderungen.
presto-coordinator	0.125	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.125	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.14.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	1.5.2	Spark-Befehlszeilen-Clients.
spark-history-server	1.5.2	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	1.5.2	In-Memory-Ausführungs-Engine für YARN.

Komponente	Version	Beschreibung
spark-yarn-slave	1.5.2	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
webserver	2.4	Apache HTTP-Server.
zeppelin-server	0.5.5-incubating-amzn-0	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.

## 4.2.0-Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

### emr-4.2.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der <code>log4j.properties</code> -Datei in Hadoop.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.



Klassifizierungen	Beschreibung
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j	Ändern Sie die Werte in der Datei 4j.properties von hive-exec-log Hive.
hive-log4j	Ändert die Werte in der hive-log4j.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.

Klassifizierungen	Beschreibung
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
presto-connector-hive	Ändert die Werte in der hive.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
spark-metrics	Ändert die Werte in der metrics.properties-Datei in Spark.

Klassifizierungen	Beschreibung
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.

## Amazon-EMR-Version 4.1.0

### 4.1.0-Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Pig](#), [Presto-Sandbox](#), [Spark](#) und [Zeppelin-Sandbox](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

### Informationen zur Anwendungsversion

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
AWS SDK for Java	1.10.27	1.10.27	Nicht nachverfolgt	Nicht nachverfolgt
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
Scala	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.6.0	-	-
HBase	-	-	-	-
HCatalog	-	-	-	-
Hadoop	2.7.1	2.6.0	2.6.0	2.6.0
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	-
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.11.0	0.11.0	0.11.0	0.10.0
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.0.1	-

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
Phoenix	-	-	-	-
Pig	0,14,0	0,14,0	0,14,0	0,14,0
Presto	-	-	-	-
Presto-Sandbox	0.130	0.125	0.119	-
Spark	1.6.0	1.5.2	1.5.0	1.4.1
Sqoop	-	-	-	-
Sqoop-Sandbox	-	-	-	-
TensorFlow	-	-	-	-
Tez	-	-	-	-
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.5.5	0,5,5	0.6.0-SNA PSHOT	-
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	-	-	-	-

## Versionshinweise zu 4.1.0

### 4.1.0-Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der

Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-amzn-*EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>emr-ddb</code>	3.0.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.0.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.1.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-s3-dist-cp</code>	2.0.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
<code>emrfs</code>	2.1.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>hadoop-client</code>	2.6.0-amzn-1	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
<code>hadoop-hdfs-datanode</code>	2.6.0-amzn-1	HDFS-Service auf Knotenebene zum Speichern von Blöcken.

Komponente	Version	Beschreibung
hadoop-hdfs-library	2.6.0-amzn-1	HDFS-Client und -Bibliothek für die Befehlszeile
hadoop-hdfs-namenode	2.6.0-amzn-1	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httfs-server	2.6.0-amzn-1	HTTP-Endpunkt für HDFS-Operationen.
hadoop-kms-server	2.6.0-amzn-1	Kryptografischer Schlüsselverwaltungsserver, der auf der Hadoop-API basiert. KeyProvider
hadoop-mapred	2.6.0-amzn-1	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.6.0-amzn-1	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.6.0-amzn-1	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hive-client	1.0.0-amzn-1	Hive-Befehlszeilen-Client.

Komponente	Version	Beschreibung
hive-metastore-server	1.0.0-amzn-1	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server	1.0.0-amzn-1	Service zur Annahme von Hive-Abfragen als Webanfragen.
hue-server	3.7.1-amzn-4	Webanwendung für die Analyse von Daten mithilfe von Hadoop-Anwendungen.
mahout-client	0.11.0	Bibliothek für Machine Learning.
mysql-server	5.5	MySQL-Datenbankserver.
oozie-client	4.0.1	Oozie-Befehlszeilen-Client.
oozie-server	4.0.1	Service für die Annahme von Oozie Workflow-Anforderungen.
presto-coordinator	0.119	Service zur Annahme von Abfragen und die Verwaltung der Abfrageausführung der Presto-Worker.
presto-worker	0.119	Service für das Ausführen von Teilen einer Abfrage.
pig-client	0.14.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	1.5.0	Spark-Befehlszeilen-Clients.



Komponente	Version	Beschreibung
spark-history-server	1.5.0	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	1.5.0	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	1.5.0	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.
zeppelin-server	0.6.0-incubating-SNAPSHOT	Webbasiertes Notizbuch, das interaktive Datenanalysen ermöglicht.

#### 4.1.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

##### emr-4.1.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.

Klassifizierungen	Beschreibung
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hdfs-encryption-zones	Konfiguriert die HDFS-Verschlüsselungszonen.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j	Ändern Sie die Werte in der Datei 4j.properties von hive-exec-log Hive.
hive-log4j	Ändert die Werte in der hive-log4j.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
hue-ini	Ändert die Werte in der INI-Datei in Hue.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
hadoop-kms-acls	Ändert die Werte in der kms-acls.xml-Datei in Hadoop.
hadoop-kms-env	Ändert die Werte in der KMS-Umgebung in Hadoop.
hadoop-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei in Hadoop.
hadoop-kms-site	Ändert die Werte in der kms-site.xml-Datei in Hadoop.

Klassifizierungen	Beschreibung
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
oozie-env	Ändert die Werte in der Oozie-Umgebung.
oozie-log4j	Ändert die Werte in der oozie-log4j.properties-Datei in Oozie.
oozie-site	Ändert die Werte in der oozie-site.xml-Datei in Oozie.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.
presto-log	Ändert die Werte in der log.properties-Datei in Presto.
presto-config	Ändert die Werte in der config.properties-Datei in Presto.
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.

Klassifizierungen	Beschreibung
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.
zeppelin-env	Ändert die Werte in der Zeppelin-Umgebung.

## Amazon-EMR-Version 4.0.0

### 4.0.0-Anwendungsversionen

Die folgenden Anwendungen werden in dieser Version unterstützt: [Hadoop](#), [Hive](#), [Mahout](#), [Pig](#) und [Spark](#).

In der folgenden Tabelle sind die in dieser Version von Amazon EMR verfügbaren Anwendungsversionen und die Anwendungsversionen der vorherigen drei Amazon-EMR-Versionen (sofern zutreffend) aufgeführt.

Einen umfassenden Verlauf der Anwendungsversionen für jede Version von Amazon EMR finden Sie in den folgenden Themen:

- [Anwendungsversionen in Amazon-EMR-7.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-6.x-Versionen](#)
- [Anwendungsversionen in Amazon-EMR-5.x-Versionen \(PNG\)](#)
- [Anwendungsversionen in Amazon-EMR-4.x-Versionen](#)

### Informationen zur Anwendungsversion

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
AWS SDK for Java	1.10.27	1.10.27	Nicht nachverfolgt	Nicht nachverfolgt
Python	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
Scala	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt	Nicht nachverfolgt
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.6.0	-	-
HBase	-	-	-	-
HCatalog	-	-	-	-
Hadoop	2.7.1	2.6.0	2.6.0	2.6.0
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	-
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.11.0	0.11.0	0.11.0	0.10.0
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.0.1	-

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
Phoenix	-	-	-	-
Pig	0,14,0	0,14,0	0,14,0	0,14,0
Presto	-	-	-	-
Presto-Sandbox	0.130	0.125	0.119	-
Spark	1.6.0	1.5.2	1.5.0	1.4.1
Sqoop	-	-	-	-
Sqoop-Sandbox	-	-	-	-
TensorFlow	-	-	-	-
Tez	-	-	-	-
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.5.5	0,5,5	0.6.0-SNA PSHOT	-
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	-	-	-	-

## Versionshinweise zu 4.0.0

### 4.0.0-Komponentenversionen

Die Komponenten, die Amazon EMR mit dieser Version installiert, sind nachstehend aufgeführt. Einige werden als Teil von Big-Data-Anwendungspaketen installiert. Andere sind nur für Amazon EMR verfügbar und werden für Systemprozesse und -Features installiert. Diese beginnen in der

Regel mit `emr` oder `aws`. Big-Data-Anwendungspakete in der aktuellsten Amazon-EMR-Version sind in der Regel die aktuelle Version, die in der Community zu finden ist. Wir stellen Community-Versionen in Amazon EMR so schnell wie möglich zur Verfügung.

Einige Komponenten in Amazon EMR unterscheiden sich von Community-Versionen. Diese Komponenten verfügen über eine Versionsbezeichnung in der Form *CommunityVersion*-amzn-*EmrVersion*. Der *EmrVersion* beginnt bei 0. Wenn zum Beispiel eine Open-Source-Community-Komponente mit dem Namen `myapp-component` der Version 2.2 dreimal für die Aufnahme in verschiedene Amazon-EMR-Versionen geändert wurde, wird ihre Version als `2.2-amzn-2` aufgeführt.

Komponente	Version	Beschreibung
<code>emr-ddb</code>	3.0.0	Amazon DynamoDB-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-goodies</code>	2.0.0	Praktische Bibliotheken für das Hadoop-Ökosystem.
<code>emr-kinesis</code>	3.0.0	Amazon Kinesis-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>emr-s3-dist-cp</code>	2.0.0	Verteilte Kopieranwendung, die für Amazon S3 optimiert ist.
<code>emrfs</code>	2.0.0	Amazon S3-Connector für Anwendungen aus dem Hadoop-Ökosystem.
<code>hadoop-client</code>	2.6.0-amzn-0	Hadoop-Befehlszeilen-Clients wie z. B. "hdfs", "Hadoop" oder "Garn".
<code>hadoop-hdfs-datanode</code>	2.6.0-amzn-0	HDFS-Service auf Knotenebene zum Speichern von Blöcken.

Komponente	Version	Beschreibung
hadoop-hdfs-namenode	2.6.0-amzn-0	HDFS-Service für die Nachverfolgung von Dateinamen und Block-Speicherorten.
hadoop-httfs-server	2.6.0-amzn-0	HTTP-Endpunkt für HDFS-Operationen.
hadoop-mapred	2.6.0-amzn-0	MapReduce Execution Engine-Bibliotheken zum Ausführen einer MapReduce Anwendung.
hadoop-yarn-nodemanager	2.6.0-amzn-0	YARN-Service für die Verwaltung von Containern auf einem einzelnen Knoten.
hadoop-yarn-resourcemanager	2.6.0-amzn-0	YARN-Service für Zuweisung und Verwaltung von Cluster-Ressourcen und verteilten Anwendungen.
hive-client	1.0.0-amzn-0	Hive-Befehlszeilen-Client.
hive-metastore-server	1.0.0-amzn-0	Service für den Zugriff auf den Hive-Metastore (ein semantisches Repository für die Speicherung von Metadaten für SQL zu Hadoop-Operationen).
hive-server	1.0.0-amzn-0	Service zur Annahme von Hive-Abfragen als Webanfragen.
mahout-client	0.10.0	Bibliothek für Machine Learning.



Komponente	Version	Beschreibung
mysql-server	5.5	MySQL-Datenbankserver.
pig-client	0.14.0-amzn-0	Pig-Befehlszeilen-Client.
spark-client	1.4.1	Spark-Befehlszeilen-Clients.
spark-history-server	1.4.1	Web-Benutzeroberfläche zum Anzeigen von protokollierten Ereignissen für die gesamte Lebensdauer einer abgeschlossenen Spark-Anwendung.
spark-on-yarn	1.4.1	In-Memory-Ausführungs-Engine für YARN.
spark-yarn-slave	1.4.1	Apache Spark-Bibliotheken, die von YARN-Slaves benötigt werden.

#### 4.0.0 Konfigurationsklassifizierungen

Mithilfe von Konfigurationsklassifizierungen können Sie Anwendungen anpassen. Diese entsprechen häufig einer XML-Konfigurationsdatei für die Anwendung, z. B. `hive-site.xml`. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

##### emr-4.0.0-Klassifizierungen

Klassifizierungen	Beschreibung
capacity-scheduler	Ändert die Werte in der <code>capacity-scheduler.xml</code> -Datei in Hadoop.
core-site	Ändert die Werte in der <code>core-site.xml</code> -Datei in Hadoop.
emrfs-site	Ändert die EMRFS-Einstellungen.

Klassifizierungen	Beschreibung
hadoop-env	Ändert die Werte in der Hadoop-Umgebung für alle Hadoop-Komponenten.
hadoop-log4j	Ändert die Werte in der log4j.properties-Datei in Hadoop.
hdfs-site	Ändert die Werte in der hdfs-site.xml-Datei in HDFS.
hive-env	Ändert die Werte in der Hive-Umgebung.
hive-exec-log4j	Ändern Sie die Werte in der Datei 4j.properties von hive-exec-log Hive.
hive-log4j	Ändert die Werte in der hive-log4j.properties-Datei in Hive.
hive-site	Ändert die Werte in der hive-site.xml-Datei in Hive.
httpfs-env	Ändert die Werte in der HTTPFS-Umgebung.
httpfs-site	Ändert die Werte in der httpfs-site.xml-Datei in Hadoop.
mapred-env	Ändern Sie die Werte in der MapReduce Anwendungsumgebung.
mapred-site	Ändern Sie die Werte in der Datei mapred-site.xml der MapReduce Anwendung.
pig-properties	Ändert die Werte in der pig.properties-Datei in Pig.
pig-log4j	Ändert die Werte in der log4j.properties-Datei in Pig.

Klassifizierungen	Beschreibung
spark	Hierbei handelt es sich um von Amazon EMR zusammengestellte Einstellungen für Apache Spark.
spark-defaults	Ändert die Werte in der spark-defaults.conf-Datei in Spark.
spark-env	Ändert die Werte in der Spark-Umgebung.
spark-log4j	Ändert die Werte in der log4j.properties-Datei in Spark.
yarn-env	Ändert die Werte in der YARN-Umgebung.
yarn-site	Ändert die Werte in der yarn-site.xml-Datei in YARN.

## AMI-Versionen 2.x und 3.x von Amazon EMR

### Note

AWS aktualisiert die TLS-Konfiguration für alle AWS API-Endpunkte auf eine Mindestversion von TLS 1.2. Amazon-EMR-Versionen 3.10 und niedriger unterstützen nur TLS-1.0/1.1-Verbindungen. Nach dem 4. Dezember 2023 können Sie mit Amazon EMR 3.10 und niedriger keine Cluster mehr erstellen.

Wenn Sie Amazon EMR 3.10 oder niedriger verwenden, empfehlen wir Ihnen, Ihre Workloads sofort zu testen und auf die neueste Amazon-EMR-Version zu migrieren. Weitere Informationen erhalten Sie im [AWS -Sicherheitsblog](#).

Die Amazon-EMR-Versionen 2.x und 3.x die auch als AMI-Versionen bezeichnet werden, stehen für bereits bestehende Lösungen zur Verfügung, die diese aus Kompatibilitätsgründen benötigen. Sie sollten mit diesen Versionen keine neuen Cluster oder Lösungen erstellen. In diesen Versionen fehlen Funktionen, die in neueren Versionen vorhanden sind. Zudem enthalten sie veraltete Anwendungspakete.

Wir empfehlen, dass Sie für das Erstellen von Lösungen die neueste Amazon-EMR-Version verwenden.

Die Unterschiede zwischen den Versionen 2.x und 3.x und den neuesten Amazon-EMR-Versionen sind signifikant. Der Umfang der Unterschiede reicht von der Erstellung und Konfiguration eines Clusters bis hin zur Port- und Verzeichnisstruktur von Anwendungen auf dem Cluster.

Dieser Abschnitt deckt die wichtigsten Amazon-EMR-Unterschiede sowie spezielle Unterschiede bei der Anwendungskonfiguration und -verwaltung ab. Die Angaben sind nicht vollständig. Wenn Sie Cluster mit den Versionen 2.x oder 3.x erstellen und nutzen, stoßen Sie möglicherweise auf Unterschiede, die in diesem Abschnitt nicht aufgeführt sind.

## Themen

- [Erstellen eines Clusters mit früheren AMI-Versionen von Amazon EMR](#)
- [Installieren von Anwendungen mit früheren AMI-Versionen von Amazon EMR](#)
- [Anpassen von Cluster- und Anwendungskonfiguration mit früheren AMI-Versionen von Amazon EMR](#)
- [Hive-Anwendungsmerkmale bei früheren AMI-Versionen von Amazon EMR](#)
- [HBase-Anwendungsmerkmale bei früheren AMI-Versionen von Amazon EMR](#)
- [Pig-Anwendungsmerkmale bei früheren AMI-Versionen von Amazon EMR](#)
- [Spark-Anwendungsmerkmale bei früheren AMI-Versionen von Amazon EMR](#)
- [Unterschiede zwischen dem DistCp S3-Hilfsprogramm und früheren AMI-Versionen von Amazon EMR](#)

## Erstellen eines Clusters mit früheren AMI-Versionen von Amazon EMR

Die Versionen 2.x und 3.x von Amazon EMR werden als AMI-Versionen bezeichnet. Ab Amazon-EMR-Version 4.0.0 und höher wird auf Versionen mit der Versionsbezeichnung verwiesen, beispielsweise `emr-5.11.0`. Diese Änderung ist am deutlichsten, wenn Sie einen Cluster programmgesteuert mithilfe von AWS CLI oder erstellen.

Wenn Sie den verwenden AWS CLI , um einen Cluster mit einer AMI-Release-Version zu erstellen, verwenden Sie die `--ami-version` Option, zum Beispiel `--ami-version 3.11.0`. Viele Optionen, Features und Anwendungen, die mit Amazon EMR 4.0.0 und höher eingeführt wurden, sind nicht verfügbar, wenn Sie `--ami-version` angeben. Weitere Informationen finden Sie unter [create-cluster](#) in der AWS CLI -Befehlsreferenz.

Der folgende AWS CLI Beispielbefehl startet einen Cluster mit einer AMI-Version.

### Note

Linux-Zeilenfortsetzungszeichen (\) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (^).

```
aws emr create-cluster --name "Test cluster" --ami-version 3.11.0 \
--applications Name=Hue Name=Hive Name=Pig \
--use-default-roles --ec2-attributes KeyName=myKey \
--instance-groups InstanceGroupType=MASTER,InstanceCount=1,\
InstanceType=m3.xlarge InstanceGroupType=CORE,InstanceCount=2,\
InstanceType=m3.xlarge --bootstrap-actions Path=s3://elasticmapreduce/bootstrap-
actions/configure-hadoop,\
Name="Configuring infinite JVM reuse",Args=["-m","mapred.job.reuse.jvm.num.tasks=-1"]
```

Programmtechnisch verwenden alle Amazon-EMR-Versionen die Aktion `RunJobFlowRequest` in der EMR-API zum Erstellen von Clustern. Der folgende Java-Beispielcode erstellt einen Cluster mit der AMI-Version 3.11.0.

```
RunJobFlowRequest request = new RunJobFlowRequest()
    .withName("AmiVersion Cluster")
    .withAmiVersion("3.11.0")
    .withInstances(new JobFlowInstancesConfig()
        .withEc2KeyName("myKeyPair")
        .withInstanceCount(1)
        .withKeepJobFlowAliveWhenNoSteps(true)
        .withMasterInstanceType("m3.xlarge")
        .withSlaveInstanceType("m3.xlarge"));
```

Der folgende `RunJobFlowRequest`-Aufruf verwendet stattdessen eine Versionsbezeichnung:

```
RunJobFlowRequest request = new RunJobFlowRequest()
    .withName("ReleaseLabel Cluster")
    .withReleaseLabel("emr-7.1.0")
    .withInstances(new JobFlowInstancesConfig()
        .withEc2KeyName("myKeyPair")
        .withInstanceCount(1)
```

```
.withKeepJobFlowAliveWhenNoSteps(true)
.withMasterInstanceType("m3.xlarge")
.withSlaveInstanceType("m3.xlarge");
```

## Konfigurieren der Cluster-Größe

Wenn Ihr Cluster ausgeführt wird, bestimmt Hadoop die Anzahl der zum Verarbeiten der Daten erforderlichen Mapper- und Reducer-Tasks. Große Cluster sollten für eine bessere Ressourcennutzung und für kürzere Verarbeitungszeiten über mehr Tasks verfügen. In der Regel behält ein EMR-Cluster seine Größe während seiner gesamten Lebensdauer. Sie legen die Anzahl der Tasks beim Erstellen des Clusters fest. Wenn Sie die Größe eines laufenden Clusters ändern, können Sie die Verarbeitung während der Cluster-Ausführung anpassen. Statt mit einer festen Anzahl von Tasks zu arbeiten, können Sie die Anzahl der Tasks während der Nutzungsdauer des Clusters anpassen. Es gibt zwei Konfigurationsoptionen, mit denen die ideale Anzahl von Tasks festgelegt werden kann:

- `mapred.map.tasksperslot`
- `mapred.reduce.tasksperslot`

Sie können beide Optionen über die `mapred-conf.xml`-Datei festlegen. Wenn Sie einen Auftrag an den Cluster senden, überprüft der Auftrags-Client die Gesamtanzahl der aktuell im gesamten Cluster verfügbaren Map- und Reduce-Slots. Der Auftrags-Client verwendet die folgenden Gleichungen, um die Anzahl der Tasks festzulegen:

- `mapred.map.tasks = mapred.map.tasksperslot * Map-Slots im Cluster`
- `mapred.reduce.tasks = mapred.reduce.tasksperslot * Reduce-Slots im Cluster`

Wenn die Anzahl der Tasks nicht konfiguriert ist, liest der Auftrags-Client nur den `tasksperslot`-Parameter aus. Sie können die Anzahl der Tasks jederzeit überschreiben – entweder über eine Bootstrap-Aktion für alle Cluster, oder einzeln pro Auftrag durch das Hinzufügen eines Schritts zum Ändern der Konfiguration.

Amazon EMR kann mit Ausfällen von Aufgabenknoten umgehen und führt den Cluster auch dann weiter aus, wenn ein Aufgabenknoten nicht mehr verfügbar ist. Amazon EMR stellt automatisch zusätzliche Aufgabenknoten als Ersatz für ausgefallene Knoten bereit.

Sie können für jeden Cluster-Schritt eine andere Anzahl von Aufgabenknoten nutzen. Sie können außerdem in einem ausgeführten Cluster einen Schritt hinzufügen, um die Anzahl der

Aufgabenknoten zu ändern. Da alle Schritte standardmäßig sequenziell ausgeführt werden, können Sie die Anzahl der ausgeführten Aufgabenknoten für einen Schritt festlegen.

## Installieren von Anwendungen mit früheren AMI-Versionen von Amazon EMR

Bei Verwendung einer AMI-Version werden Anwendungen auf verschiedene Arten installiert, z. B. mithilfe des `NewSupportedProducts` Parameters für die [RunJobFlow](#)-Aktion, mithilfe von Bootstrap-Aktionen und mithilfe der [Step-Aktion](#).

## Anpassen von Cluster- und Anwendungskonfiguration mit früheren AMI-Versionen von Amazon EMR

Mit Amazon-EMR-Version 4.0.0 wurde eine vereinfachte Methode zur Konfiguration von Anwendungen mithilfe von Konfigurationsklassifizierungen eingeführt. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#). Wenn Sie eine AMI-Version verwenden, konfigurieren Sie Anwendungen mit Bootstrap-Aktionen und Argumenten, die Sie übergeben. Beispielsweise legen die `configure-hadoop`- und `configure-daemons`-Bootstrap-Aktionen Hadoop- und YARN-spezifische Umgebungseigenschaften wie `--namenode-heap-size` fest. In neueren Versionen werden diese mit den Konfigurationsklassifizierungen `hadoop-env` und `yarn-env` konfiguriert. Bootstrap-Aktionen, die allgemeine Konfigurationen durchführen, finden Sie im [emr-bootstrap-actions Repository](#) auf Github.

Die folgenden Tabellen zeigen die Zuordnung von Bootstrap-Aktionen zu Konfigurationsklassifizierungen in neueren Amazon-EMR-Versionen.

### Hadoop

Dateiname der betroffenen Anwendung	Bootstrap-Aktion für AMI-Version	Konfigurationsklassifizierung
<code>core-site.xml</code>	<code>configure-hadoop -c</code>	<code>core-site</code>
<code>log4j.properties</code>	<code>configure-hadoop -l</code>	<code>hadoop-log4j</code>
<code>hdfs-site.xml</code>	<code>configure-hadoop -s</code>	<code>hdfs-site</code>
–	–	<code>hdfs-encryption-zones</code>

Dateiname der betroffenen Anwendung	Bootstrap-Aktion für AMI-Version	Konfigurationsklassifizierung
mapred-site.xml	configure-hadoop -m	mapred-site
yarn-site.xml	configure-hadoop -y	yarn-site
httpfs-site.xml	configure-hadoop -t	httpfs-site
capacity-scheduler.xml	configure-hadoop -z	capacity-scheduler
yarn-env.sh	configure-daemons --resource-manager-opts	yarn-env

## Hive

Dateiname der betroffenen Anwendung	Bootstrap-Aktion für AMI-Version	Konfigurationsklassifizierung
hive-env.sh	-	hive-env
hive-site.xml	hive-script --install -hive-site \${MY_HIVE_SITE_FILE}	hive-site
hive-exec-log4j.properties	-	hive-exec-log4j
hive-log4j.properties	-	hive-log4j

## EMRFS

Dateiname der betroffenen Anwendung	Bootstrap-Aktion für AMI-Version	Konfigurationsklassifizierung
emrfs-site.xml	configure-hadoop -e	emrfs-site



Dateiname der betroffenen Anwendung	Bootstrap-Aktion für AMI-Version	Konfigurationsklassifizierung
–	s3get -s s3://custom-provider.jar -d /usr/share/aws/emr/auxlib/	emrfs-site (mit neuer Einstellung fs.s3.cse.encryptionMaterialsProvider.uri )

Eine Liste aller Klassifizierungen finden Sie unter [Anwendungen konfigurieren](#).

## Variablen für die Anwendungsumgebung

Bei Verwendung einer AMI-Version wird ein `hadoop-user-env.sh`-Skript zusammen mit der Bootstrap-Aktion `configure-daemons` zur Konfiguration der Hadoop-Umgebung eingesetzt. Das Skript umfasst die folgenden Aktionen:

```
#!/bin/bash
export HADOOP_USER_CLASSPATH_FIRST=true;
echo "HADOOP_CLASSPATH=/path/to/my.jar" >> /home/hadoop/conf/hadoop-user-env.sh
```

In Amazon-EMR-Version 4.x gehen Sie genauso vor und verwenden die Konfigurationsklassifizierung `hadoop-env`, wie im folgenden Beispiel gezeigt:

```
[
  {
    "Classification": "hadoop-env",
    "Properties": {
    },
    "Configurations": [
      {
        "Classification": "export",
        "Properties": {
          "HADOOP_USER_CLASSPATH_FIRST": "true",
          "HADOOP_CLASSPATH": "/path/to/my.jar"
        }
      }
    ]
  }
]
```

Ein weiteres Beispiel verdeutlicht, dass die Verwendung von `configure-daemons` und die Übergabe von `--namenode-heap-size=2048` und `--namenode-opts=-XX:GCTimeRatio=19` den folgenden Konfigurationsklassifizierungen entspricht.

```
[
  {
    "Classification": "hadoop-env",
    "Properties": {
    },
    "Configurations": [
      {
        "Classification": "export",
        "Properties": {
          "HADOOP_DATANODE_HEAPSIZE": "2048",
          "HADOOP_NAMENODE_OPTS": "-XX:GCTimeRatio=19"
        }
      }
    ]
  }
]
```

Andere Umgebungsvariablen der Anwendung werden nicht länger in `/home/hadoop/.bashrc` definiert. Stattdessen werden sie hauptsächlich in `/etc/default`-Dateien nach Komponente oder Anwendung eingerichtet, z. B. `/etc/default/hadoop`. Wrapper-Skripts in `/usr/bin/`, die entsprechend der Anwendungs-RPMs installiert sind, können auch zusätzliche Umgebungsvariablen festlegen, ohne dass das tatsächliche Bin-Skript beteiligt werden muss.

## Service-Ports

Bei Verwendung einer AMI-Version nutzen einige Services benutzerdefinierte Ports.

### Änderungen der Port-Einstellungen

Einstellung	AMI-Version 3.x	Open-Source-Standard
<code>fs.default.name</code>	<code>hdfs://emrDeterminedIP:9000</code>	Standard ( <code>hdfs://<i>emrDeterminedIP</i> :8020</code> )
<code>dfs.datanode.address</code>	<code>0.0.0.0:9200</code>	Standard ( <code>0.0.0.0:50010</code> )
<code>dfs.datanode.http.address</code>	<code>0.0.0.0:9102</code>	Standard ( <code>0.0.0.0:50075</code> )

Einstellung	AMI-Version 3.x	Open-Source-Standard
dfs.datanode.https.address	0.0.0.0:9402	Standard (0.0.0.0:50475)
dfs.datanode.ipc.address	0.0.0.0:9201	Standard (0.0.0.0:50020)
dfs.http.address	0.0.0.0:9101	Standard (0.0.0.0:50070)
dfs.https.address	0.0.0.0:9202	Standard (0.0.0.0:50470)
dfs.secondary.http.address	0.0.0.0:9104	Standard (0.0.0.0:50090)
yarn.nodemanager.address	0.0.0.0:9103	Standard (\${yarn.nodemanager.hostname}:0)
yarn.nodemanager.localizer.address	0.0.0.0:9033	Standard (\${yarn.nodemanager.hostname}:8040)
yarn.nodemanager.webapp.address	0.0.0.0:9035	Standard (\${yarn.nodemanager.hostname}:8042)
yarn.resourcemanager.address	<i>emrDeterminedIP</i> :9022	Standard (\${yarn.resourcemanager.hostname}:8032)
yarn.resourcemanager.admin.address	<i>emrDeterminedIP</i> :9025	Standard (\${yarn.resourcemanager.hostname}:8033)
yarn.resourcemanager.resource-tracker.address	<i>emrDeterminedIP</i> :9023	Standard (\${yarn.resourcemanager.hostname}:8031)
yarn.resourcemanager.scheduler.address	<i>emrDeterminedIP</i> :9024	Standard (\${yarn.resourcemanager.hostname}:8030)
yarn.resourcemanager.webapp.address	0.0.0.0:9026	Standard (\${yarn.resourcemanager.hostname}:8088)
yarn.web-proxy.address	<i>emrDeterminedIP</i> :9046	Standard (no-value)
yarn.resourcemanager.hostname	0.0.0.0 (Standard)	<i>emrDeterminedIP</i>

**Note**

Bei *emrDeterminedIP* handelt es sich um eine IP-Adresse, die von Amazon EMR generiert wird.

## Benutzer

Bei Verwendung einer AMI-Version führt der Benutzer `hadoop` alle Prozesse aus. Zudem gehören ihm alle Dateien. Ab Amazon-EMR-Version 4.0.0 und höher sind Benutzer auf Anwendungs- und Komponentenebene vorhanden.

## Installationssequenz, installierte Artefakte und Speicherorte für die Protokolldatei

Bei Verwendung einer AMI-Version werden Anwendungs-Artefakte und deren Konfigurationsverzeichnisse im Verzeichnis `/home/hadoop/application` installiert. Wenn Sie beispielsweise das Verzeichnis Hive installiert hatten, lautete das Verzeichnis `/home/hadoop/hive`. Ab Amazon-EMR-Version 4.0.0 und höher werden Anwendungs-Artefakte im Verzeichnis `/usr/lib/application` installiert. Bei Verwendung einer AMI-Version werden Protokolldateien an unterschiedlichen Speicherorten gespeichert. In der folgenden Tabelle sind diese Speicherorte aufgeführt.

### Änderungen der Protokollspeicherorte in Amazon S3

Daemon oder Anwendung	Speicherort des Verzeichnisses
instance-state	node/ <i>instance-id</i> /instance-state/
hadoop-hdfs-namenode	<i>daemons/ instance-id/ .log</i> hadoop-hadoop-namenode
hadoop-hdfs-datanode	Daemonen/ <i>Instanz-ID</i> /hadoop-hadoop-dat anode.log
Hadoop-Garn (ResourceManager)	<i>Dämonen/ Instanz-ID/</i> yarn-hadoop- resourcemanager
hadoop-yarn (Proxy-Server)	Daemonen/ <i>Instanz-ID</i> /yarn-hadoop-proxy server

Daemon oder Anwendung	Speicherort des Verzeichnisses
mapred-historyserver	daemons/ <i>instance-id</i> /
httpfs	daemons/ <i>instance-id</i> /httpfs.log
hive-server	node/ <i>instance-id</i> /hive-server/hive-server.log
hive-metastore	node/ <i>instance-id</i> /apps/hive.log
Hive CLI	node/ <i>instance-id</i> /apps/hive.log
Benutzer- und Container-Protokolle der YARN Anwendungen	task-attempts/
Mahout	N/A
Pig	N/A
spark-historyserver	–
Verlaufsdateien des MapReduce-Auftrags	jobs/

## Command Runner

Bei Verwendung einer AMI-Version werden viele Skripts oder Programme wie `/home/hadoop/contrib/streaming/hadoop-streaming.jar` nicht in der Shell-Anmeldepfadumgebung abgelegt. Deshalb müssen Sie den vollständigen Pfad angeben, wenn Sie eine JAR-Datei wie `command-runner.jar` oder `script-runner.jar` zur Skriptausführung einsetzen. `command-runner.jar` befindet sich im AMI. Sie müssen also anders als bei `script-runner.jar` nicht die vollständige URI kennen.

## Replikationsfaktor

Mit dem Replikationsfaktor können Sie konfigurieren, wann eine Hadoop-JVM gestartet werden soll. Sie können eine neue Hadoop-JVM für jede Aufgabe starten, die eine bessere Isolierung bietet, oder Sie können JVMs mit Aufgaben teilen, um einen niedrigeren Framework-Overhead zu erzielen. Wenn Sie viele kleine Dateien verarbeiten, ist es durchaus sinnvoll, die JVM oft wieder zu

verwenden, damit sich die Kosten für den Start amortisieren. Wenn jedoch die einzelnen Aufgaben viel Zeit in Anspruch nehmen oder Prozesse eine große Datenmenge verwenden, können Sie entscheiden, die JVM nicht erneut zu verwenden, um sicherzustellen, dass der gesamte JVM-Speicher für nachfolgende Aufgaben freigegeben wird. Bei Verwendung einer AMI-Version können Sie den Replikationsfaktor mit der Bootstrap-Aktion `configure-hadoop` anpassen, um die `mapred.job.reuse.jvm.num.tasks`-Eigenschaft festzulegen.

Das folgende Beispiel zeigt die Einrichtung eines JVM-Wiederverwendungsfaktors für die unbegrenzte Wiederverwendung einer JVM.

### Note

Linux-Zeilenfortsetzungszeichen (`\`) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (`^`).





```
aws emr create-cluster --name "Test cluster" --ami-version 3.11.0 \
--applications Name=Hue Name=Hive Name=Pig \
--use-default-roles --ec2-attributes KeyName=myKey \
--instance-groups InstanceGroupType=MASTER,InstanceCount=1,InstanceType=m3.xlarge \
InstanceGroupType=CORE,InstanceCount=2,InstanceType=m3.xlarge \
--bootstrap-actions Path=s3://elasticmapreduce/bootstrap-actions/configure-hadoop,\
Name="Configuring infinite JVM reuse",Args=["-m", "mapred.job.reuse.jvm.num.tasks=-1"]
```

## Hive-Anwendungsmerkmale bei früheren AMI-Versionen von Amazon EMR

### Protokolldateien

Beim Verwenden der AMI-Versionen 2.x und 3.x von Amazon EMR werden Hive-Protokolle in `/mnt/var/log/apps/` gespeichert. Zur Unterstützung gleichzeitiger Versionen von Hive bestimmt die Hive-Version, die Sie ausführen, den Protokolldateinamen, wie in der folgenden Tabelle dargestellt.

Hive-Version	Protokolldateiname
0.13.1	hive.log

Hive-Version	Protokolldateiname
	<p> <b>Note</b></p> <p>Ab dieser Version verwendet Amazon EMR den nicht versionierten Dateinamen <code>hive.log</code>. Nebenversionen verwenden denselben Protokollspeicherort wie die Hauptversion.</p>
0.11.0	<p><code>hive_0110.log</code></p> <p> <b>Note</b></p> <p>Nebenversionen von Hive 0.11.0, z. B. 0.11.0.1, nutzen denselben Speicherort für ihre Protokolldatei wie Hive 0.11.0.</p>
0.8.1	<p><code>hive_081.log</code></p> <p> <b>Note</b></p> <p>Nebenversionen von Hive 0.8.1, z. B. Hive 0.8.1.1, nutzen denselben Speicherort für ihre Protokolldatei wie Hive 0.8.1.</p>
0.7.1	<p><code>hive_07_1.log</code></p> <p> <b>Note</b></p> <p>Nebenversionen von Hive 0.7.1, z. B. Hive 0.7.1.3 und Hive 0.7.1.4, nutzen denselben Speicherort für ihre Protokolldatei wie Hive 0.7.1.</p>
0.7	<p><code>hive_07.log</code></p>
0.5	<p><code>hive_05.log</code></p>

Hive-Version	Protokolldateiname
0.4	hive.log

## Split-Eingabe-Funktionen

Verwenden Sie Folgendes zum Implementieren von Split-Eingabe-Funktionen mit Hive-Versionen niedriger als 0.13.1 (Amazon-EMR-AMI-Versionen vor 3.11.0):

```
hive> set hive.input.format=org.apache.hadoop.hive.q1.io.HiveCombineSplitsInputFormat;
hive> set mapred.min.split.size=100000000;
```

Diese Funktionalität wurde mit Hive 0.13.1 eingestellt. Verwenden Sie folgende Einstellung, um dieselbe Funktionalität wie beim Split-Eingabeformat in AMI-Version 3.11.0 von Amazon EMR zu erhalten:

```
set hive.hadoop.supports.splittable.combineinputformat=true;
```

## Thrift-Service-Ports

Thrift ist ein RPC-Framework, das ein kompaktes binäres Serialisierungsformat definiert. Es wird dazu verwendet, Datenstrukturen für spätere Analysen beizubehalten. In der Regel konfiguriert Hive die Server für den Betrieb an folgenden Ports.

Hive-Vers ion	Port-Nummer
Hive 0.13.1	10000
Hive 0.11.0	10004
Hive 0.8.1	10003
Hive 0.7.1	10002
Hive 0.7	10001



Hive-Version	Port-Nummer
Hive 0.5	10000

Weitere Informationen über Thrift-Services finden Sie unter <http://wiki.apache.org/thrift/>.

## Verwenden von Hive zum Wiederherstellen von Partitionen

Amazon EMR enthält eine Anweisung in der Hive-Abfragesprache, mit der die Partitionen einer Tabelle aus den Tabellendaten in Amazon S3 wiederhergestellt werden. Im folgenden Beispiel ist diese Anweisung dargestellt.

```
CREATE EXTERNAL TABLE (json string) raw_impression
PARTITIONED BY (dt string)
LOCATION 's3://elastic-mapreduce/samples/hive-ads/tables/impressions';
ALTER TABLE logs RECOVER PARTITIONS;
```

Die Partitionsverzeichnisse und -daten müssen sich den dem in der Tabellendefinition angegebenen Speicherort befinden und dem Hive-Standard entsprechend benannt sein: z. B. `dt=2009-01-01`.

### Note

Nach der Version Hive 0.13.1 wird diese Funktion nativ mit `msck repair table` unterstützt. Daher wird `recover partitions` nicht unterstützt. [Weitere Informationen finden Sie unter https://cwiki.apache.org/confluence/display/Hive/+DDL+LanguageManual](https://cwiki.apache.org/confluence/display/Hive/+DDL+LanguageManual)

## Übergeben einer Hive-Variable an ein Skript

Um eine Variable mit dem an einen Hive-Schritt zu übergeben AWS CLI, geben Sie den folgenden Befehl ein, ersetzen Sie *myKey* durch den Namen Ihres EC2-Schlüsselpaars und *mybucket* durch Ihren Bucket-Namen. In diesem Beispiel ist `SAMPLE` ein Variablenwert, dem der Switch `-d` vorangestellt wird. Diese Variable ist im Hive-Skript definiert als: `${SAMPLE}`.

**Note**

Linux-Zeilenfortsetzungszeichen (\) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (^).

```
aws emr create-cluster --name "Test cluster" --ami-version 3.9 \
--applications Name=Hue Name=Hive Name=Pig \
--use-default-roles --ec2-attributes KeyName=myKey \
--instance-type m3.xlarge --instance-count 3 \
--steps Type=Hive,Name="Hive Program",ActionOnFailure=CONTINUE,\
Args=[-f,s3://elasticmapreduce/samples/hive-ads/libs/response-time-stats.q,-d,\
INPUT=s3://elasticmapreduce/samples/hive-ads/tables,-d,OUTPUT=s3://mybucket/hive-ads/\
output/, \
-d,SAMPLE=s3://elasticmapreduce/samples/hive-ads/]
```

## Angeben eines externen Metastore-Speicherorts

Das folgende Verfahren zeigt, wie Sie die Standardkonfigurationswerte für den Speicherort des Hive-Metastores überschreiben und einen Cluster mit dem neu konfigurierten Metastore-Speicherort starten.

So erstellen Sie einen Metastore außerhalb des EMR-Clusters

1. Erstellen Sie eine MySQL- oder Aurora-Datenbank mit Amazon RDS.

Weitere Informationen zum Erstellen einer Amazon-RDS-Datenbank finden Sie unter [Erste Schritte mit Amazon RDS](#).

2. Ändern Sie Ihre Sicherheitsgruppen, um JDBC-Verbindungen zwischen Ihrer Datenbank und der -Master-Sicherheitsgruppe zuzulassen. ElasticMapReduce

Informationen zum Ändern Ihrer Sicherheitsgruppen für den Zugriff finden Sie unter [Amazon-RDS-Sicherheitsgruppen](#) im Amazon-RDS-Benutzerhandbuch.

3. Legen Sie die JDBC-Konfigurationswerte in `hive-site.xml` fest:


- a. Erstellen Sie eine `hive-site.xml`-Konfigurationsdatei, die folgende Informationen enthält:

```
<configuration>
```

```
<property>
  <name>javax.jdo.option.ConnectionURL</name>
  <value>jdbc:mariadb://hostname:3306/hive?createDatabaseIfNotExist=true</
value>
  <description>JDBC connect string for a JDBC metastore</description>
</property>
<property>
  <name>javax.jdo.option.ConnectionUserName</name>
  <value>hive</value>
  <description>Username to use against metastore database</description>
</property>
<property>
  <name>javax.jdo.option.ConnectionPassword</name>
  <value>password</value>
  <description>Password to use against metastore database</description>
</property>
</configuration>
```

*Hostname* ist die DNS-Adresse der Amazon-RDS-Instance, auf der die Datenbank ausgeführt wird. *Username* und *Password* sind die Anmeldeinformationen für Ihre Datenbank. Weitere Informationen zum Verbinden mit MySQL- und Aurora-Datenbank-Instances finden Sie unter [Herstellen einer Verbindung mit einer DB-Instance auf der MySQL Datenbank Engine](#) und [Herstellen von einer Verbindung zu einem Aurora-DB-Cluster](#) im Amazon-RDS-Benutzerhandbuch.

Die JDBC-Treiber werden von Amazon EMR installiert.

 Note

Die Werteigenschaft darf keine Leerzeichen oder Wagenrückläufe enthalten. Sie sollte in einer Zeile erscheinen.

- b. Speichern Sie die `hive-site.xml`-Datei in einem Speicherort in Amazon S3, z. B. unter `s3://mybucket/hive-site.xml`.
4. Erstellen Sie einen Cluster, indem Sie den Amazon-S3-Speicherort der benutzerdefinierten Datei `hive-site.xml` angeben.

Der folgende Beispielbefehl veranschaulicht einen AWS CLI Befehl, der dies tut.

**Note**

Linux-Zeilenfortsetzungszeichen (\) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (^).

```
aws emr create-cluster --name "Test cluster" --ami-version 3.10 \  
--applications Name=Hue Name=Hive Name=Pig \  
--use-default-roles --ec2-attributes KeyName=myKey \  
--instance-type m3.xlarge --instance-count 3 \  
--bootstrap-actions Name="Install Hive Site Configuration",\  
Path="s3://region.elasticmapreduce/libs/hive/hive-script",\  
Args=["--base-path", "s3://elasticmapreduce/libs/hive", "--install-hive-site",\  
"--hive-site=s3://mybucket/hive-site.xml", "--hive-versions", "latest"]
```

## Mit Hive über JDBC verbinden

Zum Herstellen einer Verbindung mit Hive über JDBC müssen Sie den JDBC-Treiber herunterladen und einen SQL-Client installieren. Das folgende Beispiel veranschaulicht die Verwendung von SQL Workbench/J zur Verbindung mit Hive über JDBC.

So laden Sie JDBC-Treiber herunter

1. Laden Sie die entsprechenden Treiber für die Hive-Versionen herunter, auf die Sie zugreifen möchten. Entpacken Sie die Treiber. Die Hive-Version unterscheidet sich je nach dem beim Erstellen eines Amazon-EMR-Clusters ausgewählten AMI.
  - [Hive 0.13.1 JDBC-Treiber: https://s3.amazonaws.com/public/jdbc\\_1.0.4.1004.zip](https://s3.amazonaws.com/public/jdbc_1.0.4.1004.zip) [amazon-odbc-jdbc-drivers AmazonHive](#)
  - Hive 0.11.0-JDBC-Treiber: <https://mvnrepository.com/artifact/org.apache.hive/hive-jdbc/0.11.0>
  - Hive 0.8.1-JDBC-Treiber: <https://mvnrepository.com/artifact/org.apache.hive/hive-jdbc/0.8.1>
2. Installieren Sie SQL Workbench/J. Weitere Informationen finden Sie unter [Installieren und Starten von SQL Workbench/J](#) im Benutzerhandbuch für SQL Workbench/J.

3. Erstellen Sie einen SSH-Tunnel zum Cluster-Master-Knoten. Der Port für die Verbindung unterscheidet sich je nach Hive-Version. Beispielbefehle für ssh-Linux-Benutzer und PuTTY-Befehle für Windows-Benutzer finden Sie in den unten stehenden Tabellen.

#### Linux-SSH-Befehle

Hive-Version	Befehl
0.13.1	<code>ssh -o ServerAliveInterval=10 -i <i>path-to-key-file</i> -N -L 10000:localhost:10000 hadoop@ <i>master-public-dns-name</i></code>
0.11.0	<code>ssh -o ServerAliveInterval=10 -i <i>path-to-key-file</i> -N -L 10004:localhost:10004 hadoop@ <i>master-public-dns-name</i></code>
0.8.1	<code>ssh -o ServerAliveInterval=10 -i <i>path-to-key-file</i> -N -L 10003:localhost:10003 hadoop@ <i>master-public-dns-name</i></code>
0.7.1	<code>ssh -o ServerAliveInterval=10 -i <i>path-to-key-file</i> -N -L 10002:localhost:10002 hadoop@ <i>master-public-dns-name</i></code>
0.7	<code>ssh -o ServerAliveInterval=10 -i <i>path-to-key-file</i> -N -L 10001:localhost:10001 hadoop@ <i>master-public-dns-name</i></code>
0.5	<code>ssh -o ServerAliveInterval=10 -i <i>path-to-key-file</i> -N -L 10000:localhost:10000 hadoop@ <i>master-public-dns-name</i></code>

#### Windows-PuTTY-Tunneleinstellungen

Hive-Version	Tunneleinstellungen
0.13.1	Quellport: 10000 Ziel <i>master-public-dns-name</i> ::10000
0.11.0	Quellport: 10004 Ziel ::10004 <i>master-public-dns-name</i>
0.8.1	Quellport: 10003 Ziel ::10003 <i>master-public-dns-name</i>

4. Fügen Sie SQL Workbench den JDBC-Treiber hinzu.

- a. Klicken Sie im Dialogfeld Select Connection Profile (Verbindungsprofil auswählen) auf Manage Drivers (Treiber verwalten).
- b. Wählen Sie das Symbol (leere Seite) Create a new entry (Neuen Eintrag erstellen) aus.
- c. Geben Sie in das Feld Name **Hive JDBC** ein.
- d. Klicken Sie bei Library (Bibliothek) auf das Symbol Select the JAR file(s) (JAR-Datei(en) auswählen).
- e. Wählen Sie JAR-Dateien aus, wie in der folgenden Tabelle gezeigt.

Hive-Treiberversion	Hinzufügen von JAR-Dateien
0.13.1	<pre>hive_metastore.jar hive_service.jar HiveJDBC3.jar libfb303-0.9.0.jar libthrift-0.9.0.jar log4j-1.2.14.jar ql.jar slf4j-api-1.5.8.jar slf4j-log4j12-1.5.8.jar TCLIServiceClient.jar</pre>
0.11.0	<pre>hadoop-core-1.0.3.jar hive-exec-0.11.0.jar hive-jdbc-0.11.0.jar hive-metastore-0.11.0.jar hive-service-0.11.0.jar libfb303-0.9.0.jar commons-logging-1.0.4.jar slf4j-api-1.6.1.jar</pre>

Hive-Treiberversion	Hinzufügen von JAR-Dateien
0.8.1	<pre>hadoop-core-0.20.205.jar hive-exec-0.8.1.jar hive-jdbc-0.8.1.jar hive-metastore-0.8.1.jar hive-service-0.8.1.jar libfb303-0.7.0.jar libthrift-0.7.0.jar log4j-1.2.15.jar slf4j-api-1.6.1.jar slf4j-log4j12-1.6.1.jar</pre>
0.7.1	<pre>hadoop-0.20-core.jar hive-exec-0.7.1.jar hive-jdbc-0.7.1.jar hive-metastore-0.7.1.jar hive-service-0.7.1.jar libfb303.jar commons-logging-1.0.4.jar slf4j-api-1.6.1.jar slf4j-log4j12-1.6.1.jar</pre>
0.7	<pre>hadoop-0.20-core.jar hive-exec-0.7.0.jar hive-jdbc-0.7.0.jar hive-metastore-0.7.0.jar hive-service-0.7.0.jar libfb303.jar commons-logging-1.0.4.jar slf4j-api-1.5.6.jar slf4j-log4j12-1.5.6.jar</pre>

Hive-Treiberversion	Hinzufügen von JAR-Dateien
0.5	<pre> hadoop-0.20-core.jar hive-exec-0.5.0.jar hive-jdbc-0.5.0.jar hive-metastore-0.5.0.jar hive-service-0.5.0.jar libfb303.jar log4j-1.2.15.jar commons-logging-1.0.4.jar </pre>

- f. Wählen Sie im Dialogfeld Please select one driver (Bitte einen Treiber auswählen) einen Treiber gemäß der folgenden Tabelle aus und klicken Sie dann auf OK.

Hive-Version	Treiber-Klassenname
0.13.1	<code>com.amazon.hive.jdbc3.HS2Driver</code>
0.11.0	<code>org.apache.hadoop.hive.jdbc.HiveDriver.jar</code>
0.8.1	<code>org.apache.hadoop.hive.jdbc.HiveDriver.jar</code>
0.7.1	<code>org.apache.hadoop.hive.jdbc.HiveDriver.jar</code>
0.7	<code>org.apache.hadoop.hive.jdbc.HiveDriver.jar</code>
0.5	<code>org.apache.hadoop.hive.jdbc.HiveDriver.jar</code>

5. Wenn Sie zum Dialogfeld Select Connection Profile (Verbindungsprofil auswählen) zurückkehren, stellen Sie sicher, dass das Dialogfeld Driver (Treiber) auf Hive JDBC festgelegt ist, und geben Sie die JDBC-Verbindungszeichenfolge im Feld URL gemäß der folgenden Tabelle ein.



Hive-Version	JDBC-Verbindungszeichenfolge
0.13.1	<code>jdbc:hive2://localhost:10000/default</code>
0.11.0	<code>jdbc:hive://localhost:10004/default</code>
0.8.1	<code>jdbc:hive://localhost:10003/default</code>

Wenn Ihr Cluster die AMI-Version 3.3.1 oder höher im Dialogfeld Select Connection Profile (Verbindungsprofil auswählen) verwendet, geben Sie **hadoop** in das Feld Username (Benutzername) ein.

## HBase-Anwendungsmerkmale bei früheren AMI-Versionen von Amazon EMR

### Unterstützte HBase-Versionen

HBase-Version	AMI-Version	AWS CLI Konfigurationsparameter	HBase-Versionsdetails
<a href="#">0.94.18</a>	3.1.0 und höher	<pre>--ami-version 3.1  --ami-version 3.2  --ami-version 3.3  --applications Name=HBase</pre>	<ul style="list-style-type: none"> <li>Fehlerbehebungen und Verbesserungen</li> </ul>
<a href="#">0.94.7</a>	3.0-3.0.4	<pre>--ami-version 3.0</pre>	

HBase-Version	AMI-Version	AWS CLI Konfigurationsparameter	HBase-Versionsdetails
<a href="#">0.92</a>	2.2 und höher	<pre>--applications Name=HBase  --ami-version 2.2 or later  --applications Name=HBase</pre>	

## HBase-Cluster-Voraussetzungen

Ein Cluster, der mit den Amazon-EMR-AMI-Versionen 2.x und 3.x erstellt wurde, sollte die folgenden Anforderungen für HBase erfüllen.

- Das AWS CLI (optional) — Um mit HBase über die Befehlszeile zu interagieren, laden Sie die neueste Version von [herunter](#) und installieren Sie sie. AWS CLI Weitere Informationen finden Sie unter [Installieren der AWS Command Line Interface](#) im AWS Command Line Interface - Benutzerhandbuch.
- Mindestens zwei Instances (optional) – Auf dem Hauptknoten des Clusters werden der HBase-Master-Server und Zookeeper und auf den Aufgabenknoten die HBase-Regionserver ausgeführt. Für eine optimale Leistung sollten HBase-Cluster auf mindestens zwei EC2-Instances ausgeführt werden. Für Evaluierungszwecke können Sie HBase aber auch auf einem einzigen Knoten ausführen.
- Cluster mit langer Ausführung – HBase läuft nur auf Clustern mit Langzeit-Ausführung. Standardmäßig werden von der Befehlszeilenschnittstelle (CLI) und der Amazon-EMR-Konsole lang laufende Cluster erstellt.
- Ein Amazon-EC2-Schlüsselpaar (empfohlen) – Zur Verwendung des Netzwerkprotokolls Secure Shell (SSH) für die Verbindung mit dem Hauptknoten und für die Ausführung von HBase-Shell-Befehlen müssen Sie beim Erstellen des Clusters ein Amazon-EC2-Schlüsselpaar verwenden.
- Die richtigen AMI- und Hadoop-Versionen – HBase-Cluster werden derzeit nur auf Hadoop 20.205 oder höher unterstützt.
- Ganglia (optional) – Um die HBase-Leistungsmetriken zu überwachen, installieren Sie beim Erstellen des Clusters Ganglia.

- Einen Amazon-S3-Bucket für Protokolle (optional) – Die Protokolle für HBase sind auf dem Hauptknoten verfügbar. Wenn Sie möchten, dass diese Protokolle nach Amazon S3 kopiert werden, geben Sie beim Erstellen des Clusters einen S3-Bucket für den Empfang der Protokolldateien an.

## Erstellen eines Clusters mit HBase

Die folgende Tabelle enthält Optionen, die verfügbar sind, wenn mithilfe der Konsole ein Cluster mit HBase mittels einer AMI-Version von Amazon EMR erstellt wird.

Feld	Aktion
Restore from backup (Wiederherstellen mithilfe der Sicherung)	Geben Sie an, ob der HBase-Cluster mit in Amazon S3 gespeicherten Daten vorab geladen werden soll.
Backup location (Speicherort der Sicherung)	Geben Sie den URI an, in dem die Sicherung in Amazon S3 gespeichert ist, von der die Wiederherstellung erfolgen soll.
Backup version (Sicherungsversion)	Geben Sie optional den Versionsnamen der zu verwendenden Sicherung unter Backup location (Speicherort der Sicherung) an. Wenn Sie in dieses Feld keinen Wert eingeben, verwendet Amazon EMR die aktuelle Sicherung unter Backup-Speicherort, um den neuen HBase-Cluster zu füllen.
Schedule regular backups (Regelmäßige Sicherungen planen)	Geben Sie an, ob automatisch inkrementelle Sicherungen eingeplant werden sollen. Die erste Sicherung ist ein vollständiges Backup, um eine Baseline für künftige inkrementelle Sicherungen zu erstellen.
Consistent backup (Konsistente Sicherung)	Geben Sie an, ob die Sicherungen konsistent sein sollen. Bei einer konsistenten Sicherung werden die Schreibvorgänge während der anfänglichen Sicherungsstufe, der knotenübergreifenden Synchronisierung, angehalten.

Feld	Aktion
	Alle auf diese Weise angehaltenen Schreibvorgänge werden in eine Warteschlange gestellt und nach Abschluss der Synchronisierung fortgesetzt.
Backup frequency (Sicherungshäufigkeit)	Die Anzahl der Tage/Stunden/Minuten zwischen den geplanten Sicherungen.
Backup location (Speicherort der Sicherung)	Der Amazon-S3-URI, in dem die Sicherungen gespeichert werden. Der Sicherungsort sollte für jeden HBase-Cluster unterschiedlich sein, um sicherzustellen, dass differenzielle Sicherungen intakt bleiben.
Backup start time (Startzeit der Sicherung)	Geben Sie an, wann die erste Sicherung durchgeführt werden soll. Sie können für dieses Feld <code>now</code> angeben. Dann wird die erste Sicherung gestartet, sobald der Cluster ausgeführt wird. Oder geben Sie ein Datum und eine Uhrzeit im <a href="#">ISO-Format</a> ein. Beispielsweise würde <code>2012-06-15T20:00Z</code> die Startzeit auf 15. Juni 2012 um 20:00 Uhr UTC festlegen.

Der folgende AWS CLI Beispielbefehl startet einen Cluster mit HBase und anderen Anwendungen:

#### Note

Linux-Zeilenfortsetzungszeichen (`\`) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (`^`).

```
aws emr create-cluster --name "Test cluster" --ami-version 3.3 \
  --applications Name=Hue Name=Hive Name=Pig Name=HBase \
  --use-default-roles --ec2-attributes KeyName=myKey \
  --instance-type c1.xlarge --instance-count 3 --termination-protected
```

Nach dem Herstellen der Verbindung zwischen dem Hive- und HBase-Cluster (wie in der vorherigen Vorgehensweise gezeigt) können Sie auf die Daten zugreifen, die auf dem HBase-Cluster gespeichert sind, indem Sie eine externe Tabelle in Hive erstellen.

Wenn Sie das folgende Beispiel von der Hive-Eingabeaufforderung ausführen, wird eine externe Tabelle erstellt, die auf Daten verweist, die in einer HBase-Tabelle namens `inputTable` gespeichert sind. Sie können dann auf `inputTable` in Hive-Anweisungen verweisen, um Daten abzufragen und zu ändern, die im HBase-Cluster gespeichert sind.

### Note

Im folgenden Beispiel wird `protobuf-java-2.4.0a.jar` in AMI 2.3.3 verwendet. Sie sollten das Beispiel jedoch an Ihre Version anpassen. Um zu prüfen, welche Version des Protocol Buffer-JAR Sie haben, führen Sie den folgenden Befehl an der Hive-Eingabeaufforderung aus: !  
`ls /home/hadoop/lib;`

```
add jar lib/emr-metrics-1.0.jar ;
      add jar lib/protobuf-java-2.4.0a.jar ;

      set hbase.zookeeper.quorum=ec2-107-21-163-157.compute-1.amazonaws.com ;

      create external table inputTable (key string, value string)
        stored by 'org.apache.hadoop.hive.hbase.HBaseStorageHandler'
        with serdeproperties ("hbase.columns.mapping" = ":key,f1:col1")
        tblproperties ("hbase.table.name" = "t1");

      select count(*) from inputTable ;
```

## Anpassen der HBase-Konfiguration

Die Standardeinstellungen sollten zwar problemlos bei den meisten Anwendungen funktionieren, dennoch haben Sie die Möglichkeit, Ihre HBase-Konfigurationseinstellungen zu ändern. Führen Sie hierzu eines der beiden Bootstrap-Aktionsskripts aus:

- `configure-hbase-daemons`— Konfiguriert die Eigenschaften der Master-, Regionserver- und Zookeeper-Daemons. Diese Eigenschaften umfassen die Heap-Größe sowie Optionen zur Weitergabe an die Java Virtual Machine (JVM) beim Starten des HBase-Daemons. Sie definieren diese Eigenschaften als Argumente in der Bootstrap-Aktion. Diese Bootstrap-Aktion ändert die Konfigurationsdatei `/home/hadoop/conf/.sh` auf dem HBase-Cluster `hbase-user-env`.
- `configure-hbase` – Konfiguriert HBase-standortspezifische Einstellungen, z. B. den Port, an den der HBase-Master gebunden werden soll, sowie die maximale Anzahl an Wiederholversuchen einer Aktion durch den CLI-Client. Sie können diese one-by-one als Argumente in der Bootstrap-Aktion

festlegen oder den Speicherort einer XML-Konfigurationsdatei in Amazon S3 angeben. Durch diese Bootstrap-Aktion wird die Konfigurationsdatei `"/home/hadoop/conf/hbase-site.xml"` auf dem HBase-Cluster geändert.

### Note

Diese Skripts können wie andere Bootstrap-Aktionen auch nur beim Erstellen des Clusters ausgeführt werden. Sie können sie nicht verwenden, um die Konfiguration eines HBase-Clusters zu ändern, der gerade ausgeführt wird.

Wenn Sie die Aktionen `configure-hbase` oder `configure-hbase-daemonsbootstrap` ausführen, überschreiben die von Ihnen angegebenen Werte die Standardwerte. Alle Werte, die Sie nicht explizit angeben, erhalten die Standardwerte.

Die Konfiguration von HBase mit diesen Bootstrap-Aktionen ähnelt der Verwendung von Bootstrap-Aktionen in Amazon EMR zur Konfiguration von Hadoop-Einstellungen und Hadoop-Daemon-Eigenschaften. Der Unterschied besteht darin, dass HBase nicht über Speicheroptionen für einzelne Prozesse verfügt. Die Speicheroptionen werden stattdessen mit dem Argument `--daemon-opts` festgelegt, wobei *daemon* durch den Namen des zu konfigurierenden Daemons ersetzt wird.

### HBase-Daemons konfigurieren

Amazon EMR bietet eine Bootstrap-Aktion, `s3://region.elasticmapreduce/bootstrap-actions/configure-hbase-daemons`, mit der Sie die Konfiguration von HBase-Daemons ändern können. Dabei ist *region* die Region, in die Sie Ihren HBase-Cluster starten.

Um HBase-Daemons mit dem zu konfigurieren AWS CLI, fügen Sie beim Starten des Clusters die `configure-hbase-daemons` Bootstrap-Aktion hinzu, um einen oder mehrere HBase-Daemons zu konfigurieren. Sie können die folgenden Eigenschaften festlegen:

Property (Eigenschaft)	Description (Beschreibung)
<code>hbase-master-opts</code>	Optionen, die steuern, wie die JVM den Master-Daemon ausführt. Wenn diese Optionen definiert sind, werden die <code>HBASE_MASTER_OPTS</code> -Standardvariablen damit überschrieben.

Property (Eigenschaft)	Description (Beschreibung)
<code>regionserver-opts</code>	Optionen, die steuern, wie die JVM den Regionsserver-Daemon ausführt. Wenn diese Optionen definiert sind, werden die <code>HBASE_REGIONSERVER_OPTS</code> -Standardvariablen damit überschrieben.
<code>zookeeper-opts</code>	Optionen, die steuern, wie die JVM den Zookeeper-Daemon ausführt. Wenn diese Optionen definiert sind, werden die <code>HBASE_ZOOKEEPER_OPTS</code> -Standardvariablen damit überschrieben.

Weitere Informationen zu diesen Optionen finden Sie unter [hbase-env.sh](#) in der HBase-Dokumentation.

Eine Bootstrap-Aktion zum Konfigurieren von Werten für `zookeeper-opts` und `hbase-master-opts` wird im folgenden Beispiel dargestellt.

#### Note

Linux-Zeilenfortsetzungszeichen (`\`) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (`^`).

```
aws emr create-cluster --name "Test cluster" --ami-version 3.3 \
--applications Name=Hue Name=Hive Name=Pig Name=HBase \
--use-default-roles --ec2-attributes KeyName=myKey \
--instance-type c1.xlarge --instance-count 3 --termination-protected \
--bootstrap-actions Path=s3://elasticmapreduce/bootstrap-actions/configure-hbase-
daemons,\
Args=["--hbase-zookeeper-opts=-Xmx1024m -XX:GCTimeRatio=19", "--hbase-master-opts=-
Xmx2048m", "--hbase-regionserver-opts=-Xmx4096m"]
```

## HBase-Standorteinstellungen konfigurieren

Amazon EMR bietet eine Bootstrap-Aktion, `s3://elasticmapreduce/bootstrap-actions/configure-hbase`, mit der Sie die Konfiguration von HBase ändern können. Sie

können Konfigurationswerte one-by-one als Argumente in der Bootstrap-Aktion festlegen oder den Speicherort einer XML-Konfigurationsdatei in Amazon S3 angeben. Das Festlegen von Konfigurationswerten one-by-one ist nützlich, wenn Sie nur einige Konfigurationseinstellungen festlegen müssen. Sie mit einer XML-Datei zu definieren, ist hilfreich, wenn Sie viele Änderungen vornehmen müssen oder Ihre Konfigurationseinstellungen für die Wiederverwendung speichern möchten.

### Note

Sie können dem Amazon-S3-Bucket-Namen einen Regionspräfix voranstellen, z. B. `s3://region.elasticmapreduce/bootstrap-actions/configure-hbase`, wobei *region* die Region ist, in die Sie Ihren HBase-Cluster starten.

Durch diese Bootstrap-Aktion wird die Konfigurationsdatei `/home/hadoop/conf/hbase-site.xml` auf dem HBase-Cluster geändert. Die Bootstrap-Aktion kann nur beim Starten des HBase-Clusters ausgeführt werden.

Weitere Informationen zu den HBase-Standorteinstellungen, die Sie konfigurieren können, finden Sie unter [Standardkonfiguration](#) in der HBase-Dokumentation.

Legen Sie die Bootstrap-Aktion `configure-hbase` beim Starten des HBase-Clusters fest und geben Sie die zu ändernden Werte in `hbase-site.xml` an.

Um einzelne HBase-Site-Einstellungen mit dem zu spezifizieren AWS CLI

- Um die Einstellung `hbase.hregion.max.filesize` zu ändern, geben Sie den folgenden Befehl ein und ersetzen Sie *myKey* durch den Namen Ihres Amazon-EC2-Schlüsselpaares.

### Note

Linux-Zeilenfortsetzungszeichen (`\`) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (`^`).

```
aws emr create-cluster --name "Test cluster" --ami-version 3.3 \  
--applications Name=Hue Name=Hive Name=Pig Name=HBase \  
--use-default-roles --ec2-attributes KeyName=myKey \  

```



```
--instance-type c1.xlarge --instance-count 3 --termination-protected \  
--bootstrap-actions Path=s3://elasticmapreduce/bootstrap-actions/configure-  
hbase,Args=["-s", "hbase.hregion.max.filesize=52428800"]
```

Um HBase-Site-Einstellungen mit einer XML-Datei anzugeben, verwenden Sie den AWS CLI

1. Erstellen Sie eine benutzerdefinierte Version von `hbase-site.xml`. Die benutzerdefinierte Datei muss eine gültige XML-Datei sein. Um Fehlerquellen zu minimieren, starten Sie mit der Standardkopie von `hbase-site.xml` auf dem HBase-Hauptknoten von Amazon EMR unter `/home/hadoop/conf/hbase-site.xml`. Bearbeiten Sie eine Kopie dieser Datei, anstatt eine Datei von Grund auf neu zu erstellen. Sie können Ihrer neuen Datei einen neuen Namen geben oder den Dateinamen `hbase-site.xml` belassen.
2. Laden Sie die benutzerdefinierte Datei `hbase-site.xml` in einen Amazon-S3-Bucket hoch. Es sollten die Berechtigungen so eingerichtet sein, dass das AWS Konto, das den Cluster startet, auf die Datei zugreifen kann. Wenn das AWS Konto, das den Cluster startet, auch den Amazon S3 S3-Bucket besitzt, hat es Zugriff.
3. Definieren Sie die Bootstrap-Aktion `configure-hbase` beim Starten des HBase-Clusters und fügen Sie den Speicherort Ihrer benutzerdefinierten Datei `hbase-site.xml` ein. Beim folgenden Beispiel sind die HBase-Standortkonfigurationswerte auf Werte in der Datei `s3://mybucket/my-hbase-site.xml` festgelegt. Geben Sie den folgenden Befehl ein, ersetzen Sie *myKey* durch den Namen Ihres EC2-Schlüsselpaares und *mybucket* durch den Namen Ihres Amazon-S3-Buckets.

#### Note

Linux-Zeilenfortsetzungszeichen (`\`) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (`^`).

```
aws emr create-cluster --name "Test cluster" --ami-version 3.3 \  
--applications Name=Hue Name=Hive Name=Pig Name=HBase \  
--use-default-roles --ec2-attributes KeyName=myKey \  
--instance-type c1.xlarge --instance-count 3 --termination-protected \  
--bootstrap-actions Path=s3://elasticmapreduce/bootstrap-actions/configure-  
hbase,Args=["--site-config-file", "s3://mybucket/config.xml"]
```

Wenn Sie mehr als eine Option zum Anpassen der HBase-Operation angeben, müssen Sie jedem Schlüssel-Wert-Paar einen `-s`-Optionsschalter voranstellen, wie im folgenden Beispiel gezeigt:

```
--bootstrap-actions s3://elasticmapreduce/bootstrap-actions/configure-hbase,Args=["-s","zookeeper.session.timeout=60000"]
```

Wenn der Proxy eingestellt und die SSH-Verbindung geöffnet ist, können Sie die HBase-Benutzeroberfläche aufrufen, indem Sie ein Browserfenster mit `http://:60010/master-status` öffnen *master-public-dns-name*, in dem sich die öffentliche DNS-Adresse des Master-Knotens *master-public-dns-name* im HBase-Cluster befindet.

Sie können die aktuellen HBase-Protokolle anzeigen, indem Sie sich über SSH mit dem Master-Knoten verbinden und zum Verzeichnis `mnt/var/log/hbase` navigieren. Diese Protokolle sind nicht verfügbar, nachdem der Cluster beendet wurde, es sei denn, Sie aktivieren beim Starten des Clusters die Protokollierung nach Amazon S3.

## HBase sichern und wiederherstellen

Amazon EMR bietet die Möglichkeit, Ihre HBase-Daten nach Amazon S3 zu sichern, entweder manuell oder über einen automatisierten Zeitplan. Sie können sowohl vollständige als auch inkrementelle Sicherungen durchführen. Nachdem Sie eine Version von HBase-Daten gesichert haben, können Sie diese auf einen HBase-Cluster wiederherstellen. Sie können die Wiederherstellung auf einem HBase-Cluster durchführen, der gerade ausgeführt wird, oder einen neuen Cluster starten, der mit den gesicherten Daten vorab gefüllt wird.

Während des Sicherungsvorgangs fährt HBase fort, Schreibbefehle auszuführen. Dadurch wird zwar sichergestellt, dass Ihr Cluster während der Sicherung verfügbar bleibt, es besteht aber das Risiko von Inkonsistenzen zwischen den Daten, die gesichert werden, und den Schreibvorgängen, die parallel ausgeführt werden. Um die Inkonsistenzen zu verstehen, die möglicherweise auftreten, müssen Sie berücksichtigen, dass HBase Schreibvorgänge auf die Knoten in seinem Cluster verteilt. Wenn ein Schreibvorgang erfolgt, nachdem ein bestimmter Knoten abgefragt wurde, werden diese Daten nicht in das Sicherungsarchiv aufgenommen. Möglicherweise stellen Sie sogar fest, dass frühere Schreibvorgänge auf dem HBase-Cluster (gesendet an einen Knoten, der bereits abgefragt

wurde) nicht im Sicherungsarchiv vorhanden sind, während spätere Schreibvorgänge (gesendet an einen Knoten vor dessen Abfrage) enthalten sind.

Wenn eine konsistente Sicherung erforderlich ist, müssen Sie die Schreibvorgänge nach HBase während des ersten Teils des Sicherungsprozesses, der knotenübergreifenden Synchronisierung, anhalten. Sie tun dies, indem Sie beim Anfordern einer Sicherung den Parameter `--consistent` angeben. Mit diesem Parameter werden die Schreibvorgänge während dieses Zeitraums in eine Warteschlange gestellt und ausgeführt, sobald die Synchronisierung abgeschlossen ist. Sie können auch regelmäßige Sicherungen einplanen. Dadurch werden Inkonsistenzen, die im Laufe der Zeit entstehen, aufgelöst, da die Daten, die bei einer Sicherung nicht berücksichtigt wurden, im folgenden Vorgang gesichert werden.

Wenn Sie HBase-Daten sichern, sollten Sie für jeden Cluster ein anderes Sicherungsverzeichnis angeben. Eine einfache Möglichkeit dafür ist die Verwendung der Cluster-Kennung als Bestandteil des für das Sicherungsverzeichnis angegebenen Pfads. z. B. `s3://mybucket/backups/j-3AEXXXXXX16F2`. Dadurch wird sichergestellt, dass alle künftigen inkrementellen Backups auf den richtigen HBase-Cluster verweisen.

Wenn Sie alte, nicht mehr benötigte Sicherungsdateien löschen möchten, wird empfohlen, dass Sie zuerst eine vollständige Sicherung Ihrer HBase-Daten durchführen. Auf diese Weise wird sichergestellt, dass alle Daten erhalten bleiben. Außerdem bietet Ihnen dies eine Baseline für zukünftige inkrementelle Sicherungen. Nachdem die vollständige Sicherung abgeschlossen ist, können Sie zum Sicherungsverzeichnis navigieren und die alten Sicherungsdateien manuell löschen.

Der HBase-Backup-Prozess verwendet S3 DistCp für den Kopiervorgang, der bestimmte Einschränkungen in Bezug auf den Speicherplatz für temporäre Dateien aufweist.

## HBase mit der Konsole sichern und wiederherstellen

Die Konsole bietet die Möglichkeit, einen neuen Cluster zu starten und mit Daten aus einer vorherigen HBase-Sicherung zu füllen. Außerdem können Sie periodische inkrementelle Sicherungen von HBase-Daten einplanen. Über die CLI stehen zusätzliche Sicherungs- und Wiederherstellungsfunktion zur Verfügung, z. B. können Daten auf einem bereits ausgeführten Cluster wiederhergestellt, manuelle Sicherungen durchgeführt und automatisierte vollständige Sicherungen terminiert werden.

So füllen Sie einen neuen Cluster mit archivierten HBase-Daten über die Konsole

1. Navigieren Sie zur neuen Amazon-EMR-Konsole und wählen Sie in der Seitennavigation die Option Zur alten Konsole wechseln aus. Weitere Informationen darüber, was Sie erwartet, wenn Sie zur alten Konsole wechseln, finden Sie unter [Verwenden der alten Konsole](#).
2. Wählen Sie Cluster erstellen.
3. Wählen Sie im Abschnitt Software Configuration (Softwarekonfiguration) unter Additional Applications (Zusätzliche Anwendungen) die Optionen HBase und Configure and add aus.
4. Klicken Sie auf das Dialogfeld Add Application (Anwendung hinzufügen) und aktivieren Sie das Kontrollkästchen Restore From Backup (Wiederherstellen mithilfe der Sicherung).
5. Geben Sie für Backup Location (Speicherort der Sicherung) den Speicherort der Sicherung an, die Sie in den neuen HBase-Cluster laden. Dies sollte eine Amazon-S3-URL im Format `s3://myawsbucket/backups/` sein.
6. Für Backup Version (Sicherungsversion) haben Sie durch Angabe eines Werts die Möglichkeit, den Namen einer zu ladenden Sicherungsversion anzugeben. Wenn Sie keinen Wert für Backupversion (Sicherungsversion) angeben, lädt Amazon EMR die letzte Sicherung in das angegebene Verzeichnis.
7. Wählen Sie Add (Hinzufügen) und fahren Sie fort, um den Cluster mit den anderen Optionen wie gewünscht zu erstellen.

So planen Sie automatische Sicherungen von HBase-Daten mit der Konsole

1. Wählen Sie im Abschnitt Software Configuration (Softwarekonfiguration) unter Additional Applications (Zusätzliche Anwendungen) die Optionen HBase und Configure and add aus.
2. Wählen Sie Schedule Regular Backups (Regelmäßige Sicherungen planen).
3. Geben Sie an, ob die Sicherungen konsistent sein sollen. Bei einer konsistenten Sicherung werden die Schreibvorgänge während der anfänglichen Sicherungsstufe, der knotenübergreifenden Synchronisierung, angehalten. Alle auf diese Weise angehaltenen Schreibvorgänge werden in eine Warteschlange gestellt und nach Abschluss der Synchronisierung fortgesetzt.
4. Legen Sie fest, wie oft Sicherungen durchgeführt werden sollen, indem Sie eine Zahl für Backup Frequency (Sicherungshäufigkeit) eingeben, und wählen Sie Days (Tage), Hours (Stunden) oder Minutes (Minuten) aus. Bei der ersten automatischen Sicherung handelt es sich um eine vollständige Sicherung. Danach speichert Amazon EMR inkrementelle Backups basierend auf dem von Ihnen festgelegten Zeitplan.

5. Geben Sie den Speicherort in Amazon S3 an, in dem die Sicherungen gespeichert werden sollen. Jeder HBase-Cluster sollte an einem anderen Speicherort in Amazon S3 gespeichert werden, um sicherzustellen, dass inkrementelle Backups korrekt berechnet werden.
6. Geben Sie an, wann die erste Sicherung erfolgen soll, indem Sie einen Wert für Backup Start Time (Startzeit der Sicherung) eingeben. Sie können für dieses Feld now angeben. Dann wird die erste Sicherung gestartet, sobald der Cluster ausgeführt wird. Oder geben Sie ein Datum und eine Uhrzeit im [ISO-Format](#) ein. Beispielsweise würde 2013-09-26T20:00Z die Startzeit auf 26. September 2013 um 20:00 Uhr UTC festlegen.
7. Wählen Sie Hinzufügen aus.
8. Fahren Sie mit der Erstellung des Clusters je nach Bedarf mit den anderen Optionen fort.

## Überwachen Sie HBase mit CloudWatch

Amazon EMR meldet drei Metriken CloudWatch, anhand derer Sie Ihre HBase-Backups überwachen können. Diese Messwerte werden CloudWatch in Intervallen von fünf Minuten aktualisiert und kostenlos zur Verfügung gestellt.

Metrik	Beschreibung
HBaseBackupFailed	<p>Gibt an, ob die letzte Sicherung fehlgeschlagen ist. Sie ist standardmäßig auf 0 gesetzt und wird in 1 geändert, wenn der vorhergehende Sicherungsversuch fehlgeschlagen ist. Diese Metrik wird nur für HBase-Cluster ausgewertet.</p> <p>Anwendungsfall: Überwachen von HBase-Sicherungen</p> <p>Einheiten: Anzahl</p>
HBaseMostRecentBackupDuration	<p>Dauer, bis die vorherige Sicherung abgeschlossen ist. Diese Metrik wird unabhängig davon erstellt, ob die letzte Sicherung erfolgreich war oder fehlgeschlagen ist. Während die Sicherung ausgeführt wird, gibt die Metrik die Anzahl der Minuten seit dem Start der Sicherung zurück. Diese Metrik wird nur für HBase-Cluster ausgewertet.</p>

Metrik	Beschreibung
	Anwendungsfall: Überwachen von HBase-Sicherungen Einheiten: Minuten
HBaseTimeSinceLastSuccessfulBackup	Anzahl der verstrichenen Minuten nach der letzten erfolgreichen HBase-Sicherung auf Ihrem Cluster. Diese Metrik wird nur für HBase-Cluster ausgewertet. Anwendungsfall: Überwachen von HBase-Sicherungen Einheiten: Minuten

## Konfigurieren von Ganglia für HBase

Sie konfigurieren Ganglia für HBase mithilfe der `configure-hbase-for-gangliaBootstrap`-Aktion. Diese Bootstrap-Aktion konfiguriert HBase so, dass Metriken in Ganglia veröffentlicht werden.

Sie müssen HBase und Ganglia beim Starten des Clusters konfigurieren. Die Ganglia-Auswertungen können nicht zu einem aktuell ausgeführten Cluster hinzugefügt werden.

Ganglia speichert Protokolldateien auch auf dem Server unter `/mnt/var/log/ganglia/rrds`. Wenn Sie Ihren Cluster so konfiguriert haben, dass Protokolldateien dauerhaft in einem Amazon-S3-Bucket gespeichert werden, werden die Ganglia-Protokolldateien auch dort dauerhaft abgelegt.

Um einen Cluster mit Ganglia für HBase zu starten, verwenden Sie die `configure-hbase-for-gangliaBootstrap`-Aktion, wie im folgenden Beispiel gezeigt.

### Note

Linux-Zeilenfortsetzungszeichen (`\`) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (`^`).

```
aws emr create-cluster --name "Test cluster" --ami-version 3.3 \
--applications Name=Hue Name=Hive Name=Pig Name=HBase Name=Ganglia \
--use-default-roles --ec2-attributes KeyName=myKey \
--instance-type c1.xlarge --instance-count 3 --termination-protected \
```

```
--bootstrap-actions Path=s3://elasticmapreduce/bootstrap-actions/configure-hbase-for-ganglia
```

Nachdem der Cluster gestartet wurde, für den Ganglia konfiguriert ist, können Sie auf die Ganglia-Grafiken und -Berichte über die Benutzeroberfläche auf dem Master-Knoten zugreifen.

## Pig-Anwendungsmerkmale bei früheren AMI-Versionen von Amazon EMR

### Unterstützte Pig-Versionen

Welche Pig-Version Sie zu Ihrem Cluster hinzufügen können, hängt von der verwendeten Amazon-EMR-AMI-Version und der Hadoop-Version ab. Die folgende Tabelle zeigt, welche AMI- und Hadoop-Versionen mit den verschiedenen Pig-Versionen kompatibel sind. Wir empfehlen die Verwendung der neuesten Pig-Version, um die Vorteile der Leistungsverbesserungen und neuen Funktionen zu nutzen.

Wenn Sie die API zur Installation von Pig verwenden, wird die Standardversion verwendet, sofern Sie nicht `--pig-versions` als Argument für den Schritt angeben, der Pig während des Aufrufs von in den Cluster lädt. [RunJobFlow](#)

Pig-Version	AMI-Version	Konfigurationsparameter	Details zur Pig-Version
0.12.0 <a href="#">Versionshinweise</a> <a href="#">Dokumentation</a>	3.1.0 und höher	<code>--ami-version 3.1</code> <code>--ami-version 3.2</code> <code>--ami-version 3.3</code>	Implementiert die Unterstützung der folgenden Elemente: <ul style="list-style-type: none"> <li>• Streamen von UDFs ohne JVM-Implementierungen</li> <li>• Operatoren ASSERT und IN</li> <li>• CASE-Ausdruck</li> <li>• AvroStorage als integrierte Pig-Funktion.</li> <li>• ParquetLoader und ParquetStorer</li> </ul>

Pig-Version	AMI-Version	Konfigurationsparameter	Details zur Pig-Version
			<p>als eingebaute Funktionen</p> <ul style="list-style-type: none"> <li>• BigInteger und BigDecimal Typen</li> </ul>
0.11.1.1 <a href="#">Versionshinweise</a> <a href="#">Dokumentation</a>	2.2 und höher	<pre>--pig-versions 0.11.1.1  --ami-version 2.2</pre>	Verbessert die Leistung des LOAD-Befehls PigStorage, wenn sich die Eingabe in Amazon S3 befindet.
0.11.1 <a href="#">Versionshinweise</a> <a href="#">Dokumentation</a>	2.2 und höher	<pre>--pig-versions 0.11.1  --ami-version 2.2</pre>	Integriert Unterstützung für JDK 7, Hadoop 2, benutzerdefinierte Groovy-Funktionen, SchemaTuple Optimierung, neue Operatoren und mehr. Weitere Informationen finden Sie im <a href="#">Veränderungsprotokoll für Pig 0.11.1</a> .
0.9.2.2 <a href="#">Versionshinweise</a> <a href="#">Dokumentation</a>	2.2 und höher	<pre>--pig-versions 0.9.2.2  --ami-version 2.2</pre>	Implementiert die Unterstützung von Hadoop 1.0.3.
0.9.2.1 <a href="#">Versionshinweise</a> <a href="#">Dokumentation</a>	2.2 und höher	<pre>--pig-versions 0.9.2.1  --ami-version 2.2</pre>	Implementiert die Unterstützung von MapR.



Pig-Version	AMI-Version	Konfigurationsparameter	Details zur Pig-Version
0.9.2 <a href="#">Versionshinweise</a> <a href="#">Dokumentation</a>	2.2 und höher	<code>--pig-versions 0.9.2</code> <code>--ami-version 2.2</code>	Umfasst mehrere Leistungsverbesserungen und Fehlerkorrekturen. Die vollständigen Informationen zu den Änderungen mit Pig 0.9.2 finden Sie im <a href="#">Veränderungsprotokoll für Pig 0.9.2</a> .
0.9.1 <a href="#">Versionshinweise</a> <a href="#">Dokumentation</a>	2.0	<code>--pig-versions 0.9.1</code> <code>--ami-version 2.0</code>	
0.6 <a href="#">Versionshinweise</a>	1,0	<code>--pig-versions 0.6</code> <code>--ami-version 1.0</code>	
0.3 <a href="#">Versionshinweise</a>	1,0	<code>--pig-versions 0.3</code> <code>--ami-version 1.0</code>	

## Details zur Pig-Version

Amazon EMR unterstützt bestimmte Pig-Versionen, auf die möglicherweise zusätzliche Amazon-EMR-Patches angewendet wurden. Sie können konfigurieren, welche Pig-Version auf den Amazon-EMR-Clustern ausgeführt wird. Weitere Information dazu finden Sie unter [Apache Pig](#). In den

folgenden Abschnitten werden die verschiedenen Pig-Versionen und die auf die in Amazon EMR geladenen Versionen angewendeten Patches beschrieben.

## Pig-Patches

Dieser Abschnitt beschreibt die benutzerdefinierten Patches, die auf die in Amazon EMR verfügbaren Pig-Versionen angewendet werden.

### Pig-0.11.1.1-Patches

Die Amazon EMR-Version von Pig 0.11.1.1 ist eine Wartungsversion, die die Leistung des LOAD-Befehls verbessert, PigStorage wenn sich die Eingabe in Amazon S3 befindet.

### Pig 0.11.1-Patches

Die Amazon-EMR-Version von Pig 0.11.1 enthält alle Aktualisierungen der Apache Software Foundation und die kumulativen Amazon-EMR-Patches aus Pig-Version 0.9.2.2. Es gibt jedoch keine neuen Amazon-EMR-spezifischen Patches in Pig 0.11.1.

### Pig-0.9.2-Patches

Apache Pig 0.9.2 ist eine Wartungsversion von Pig. Das Amazon-EMR-Team hat die folgenden Patches auf die Amazon-EMR-Version von Pig 0.9.2 angewendet.

Patch	Beschreibung
PIG-1429	<p>Fügt den Datentyp Boolean als Pig-Datentyp als primären Datentyp zu Pig hinzu. Weitere Informationen finden Sie unter <a href="https://issues.apache.org/jira/browse/PIG-1429">https://issues.apache.org/jira/browse/PIG-1429</a>.</p> <p>Status: Ausgeführt</p> <p>Fixed in Apache Pig Version: (In Apache Pig-Version behoben:) 0.10</p>
PIG-1824	<p>Unterstützung von Import-Module in Jython UDF. Weitere Informationen finden Sie unter <a href="https://issues.apache.org/jira/browse/PIG-1824">https://issues.apache.org/jira/browse/PIG-1824</a>.</p> <p>Status: Ausgeführt</p>

Patch	Beschreibung
	Fixed in Apache Pig Version: (In Apache Pig-Version behoben:) 0.10
PIG-2010	<p>Im Paket registrierte JARs im verteilten Cache. Weitere Informationen finden Sie unter <a href="https://issues.apache.org/jira/browse/PIG-2010">https://issues.apache.org/jira/browse/PIG-2010</a>.</p> <p>Status: Ausgeführt</p> <p>Fixed in Apache Pig Version: (In Apache Pig-Version behoben:) 0.11</p>
PIG-2456	<p>~/pigbootup-Datei hinzugefügt, über die Benutzer Standard-Pig-Anweisungen angeben können. Weitere Informationen finden Sie unter <a href="https://issues.apache.org/jira/browse/PIG-2456">https://issues.apache.org/jira/browse/PIG-2456</a>.</p> <p>Status: Ausgeführt</p> <p>Fixed in Apache Pig Version: (In Apache Pig-Version behoben:) 0.11</p>
PIG-2623	<p>Unterstützung für die Verwendung von Amazon-S3-Pfaden zur Registrierung von UDFs. Weitere Informationen finden Sie unter <a href="https://issues.apache.org/jira/browse/PIG-2623">https://issues.apache.org/jira/browse/PIG-2623</a>.</p> <p>Status: Ausgeführt</p> <p>Fixed in Apache Pig Version: (In Apache Pig-Version behoben:) 0.10, 0.11</p>

### Pig-0.9.1-Patches

Das Amazon-EMR-Team hat die folgenden Patches auf die Amazon-EMR-Version von Pig 0.9.1 angewendet.

Patch	Beschreibung
Unterstützung für JAR-Dateien und Pig-Skripts in DFS	<p>Fügt Unterstützung zur Ausführung von Skripten und zur Registrierung von JAR-Dateien, die in HDFS, Amazon S3 oder anderen verteilten Dateisystemen gespeichert sind, hinzu. Weitere Informationen finden Sie unter <a href="https://issues.apache.org/jira/browse/PIG-1505">https://issues.apache.org/jira/browse/PIG-1505</a>.</p> <p>Status: Ausgeführt</p> <p>Fixed in Apache Pig Version: (In Apache Pig-Version behoben:) 0.8.0</p>
Unterstützung mehrerer Dateisysteme in Pig	<p>Fügt Unterstützung für Pig-Skripts zum Lesen von Daten aus einem Dateisystem und Schreiben in einem anderen hinzu. Weitere Informationen finden Sie unter <a href="https://issues.apache.org/jira/browse/PIG-1564">https://issues.apache.org/jira/browse/PIG-1564</a>.</p> <p>Status: Nicht ausgeführt</p> <p>Fixed in Apache Pig Version: (In Apache Pig-Version behoben:) –</p>
Piggybank-Datetime- und String-UDFs hinzugefügt	<p>Datetime- und String-UDFs für benutzerdefinierte Pig-Skripts hinzugefügt. Weitere Informationen finden Sie unter <a href="https://issues.apache.org/jira/browse/PIG-1565">https://issues.apache.org/jira/browse/PIG-1565</a>.</p> <p>Status: Nicht ausgeführt</p> <p>Fixed in Apache Pig Version: (In Apache Pig-Version behoben:) –</p>

## Interaktive Pig-Cluster und Batch-Pig-Cluster

Amazon EMR ermöglicht die Ausführung von Pig-Skripten in zwei Modi:

- Interaktive
- Stapel

Wenn Sie einen Cluster mit langer Laufzeit über die Konsole oder die starten AWS CLI, können Sie sich mit dem Master-Knoten als Hadoop-Benutzer verbinden und ssh die Grunt-Shell verwenden, um Ihre Pig-Skripte interaktiv zu entwickeln und auszuführen. Durch die interaktive Verwendung von Pig können Sie Pig-Skripts einfacher überarbeiten als im Batch-Modus. Nachdem Sie das Pig-Skript im interaktiven Modus erfolgreich überarbeitet haben, können Sie das Skript in Amazon S3 hochladen und den Batch-Modus nutzen, um das Skript in der Produktionsumgebung auszuführen. Sie können außerdem interaktiv Pig-Befehle an einen aktuell ausgeführten Cluster senden, um Daten bei Bedarf zu analysieren und zu transformieren.

Im Batch-Modus laden Sie in einem einzigen Schritt Ihr Pig-Skript in Amazon S3 hoch und senden gleichzeitig den Auftrag an den Cluster. Pig-Schritte können an lang- oder kurzlebige Cluster und an Übergangs-Cluster gesendet werden.

## Spark-Anwendungsmerkmale bei früheren AMI-Versionen von Amazon EMR

### Spark im interaktiven Modus oder im Batch-Modus verwenden

Amazon EMR ermöglicht die Ausführung von Spark-Anwendungen in zwei Modi:

- Interactive
- Stapel

Wenn Sie einen Cluster mit langer Laufzeit über die Konsole oder die starten AWS CLI, können Sie als Hadoop-Benutzer über SSH eine Verbindung zum Master-Knoten herstellen und die Spark-Shell verwenden, um Ihre Spark-Anwendungen interaktiv zu entwickeln und auszuführen. Mit der interaktiven Verwendung von Spark ist das Entwickeln von Prototypen für Spark-Anwendungen und das Testen dieser Anwendungen einfacher als in einer Batch-Umgebung. Wenn Sie die Spark-Anwendung im interaktiven Modus erfolgreich überarbeitet haben, können Sie die JAR-Anwendungsdatei oder das Python-Programm im lokalen Dateisystem des Hauptknotens des Clusters in Amazon S3 ablegen. Anschließend können Sie die Anwendung als Batch-Workflow senden.

Laden Sie im Batch-Modus das Spark-Skript in Amazon S3 oder ins Dateisystem des lokalen Hauptknotens hoch und senden Sie dann die Arbeit als Schritt an den Cluster. Spark-Schritte können an lang- oder kurzlebige Cluster gesendet werden.

## Erstellen eines Clusters mit installiertem Spark

So starten Sie einen Cluster mit installiertem Spark über die Console

1. Navigieren Sie zur neuen Amazon-EMR-Konsole und wählen Sie in der Seitennavigation die Option Zur alten Konsole wechseln aus. Weitere Informationen darüber, was Sie erwartet, wenn Sie zur alten Konsole wechseln, finden Sie unter [Verwenden der alten Konsole](#).
2. Wählen Sie Cluster erstellen.
3. Wählen Sie für Software Configuration (Softwarekonfiguration) die benötigte AMI-Version aus.
4. Wählen Sie unter Applications to be installed (Zu installierende Anwendungen) in der Liste Spark und anschließend Configure and add (Konfigurieren und hinzufügen) aus.
5. Sie können Argumente hinzufügen, um die Spark-Konfiguration bei Bedarf zu ändern. Weitere Informationen finden Sie unter [Konfigurieren von Spark](#). Wählen Sie Hinzufügen aus.
6. Wählen Sie nach Bedarf weitere Optionen und anschließend Create cluster (Cluster erstellen) aus.

Das folgende Beispiel zeigt, wie Sie einen Cluster mit Spark mittels Java erstellen:

```
AmazonElasticMapReduceClient emr = new AmazonElasticMapReduceClient(credentials);
SupportedProductConfig sparkConfig = new SupportedProductConfig()
    .withName("Spark");

RunJobFlowRequest request = new RunJobFlowRequest()
    .withName("Spark Cluster")
    .withAmiVersion("3.11.0")
    .withNewSupportedProducts(sparkConfig)
    .withInstances(new JobFlowInstancesConfig()
        .withEc2KeyName("myKeyName")
        .withInstanceCount(1)
        .withKeepJobFlowAliveWhenNoSteps(true)
        .withMasterInstanceType("m3.xlarge")
        .withSlaveInstanceType("m3.xlarge")
    );
RunJobFlowResult result = emr.runJobFlow(request);
```

## Konfigurieren von Spark

Sie konfigurieren Spark, wenn Sie einen Cluster erstellen, indem Sie die Bootstrap-Aktion ausführen, die sich im [emr-bootstrap-actionsaws-labs/spark-Repository](#) auf Github befindet. Informationen zu den

Argumenten, die von der Bootstrap-Aktion akzeptiert werden, finden Sie in der [README-Datei](#) dieses Repositorys. Die Bootstrap-Aktion konfiguriert Eigenschaften in der Datei `$SPARK_CONF_DIR/spark-defaults.conf`. Weitere Informationen zu den Einstellungen finden Sie im Thema zur Spark-Konfiguration in der Spark-Dokumentation. Sie können „latest“ in der folgenden URL durch die Nummer der Spark-Version ersetzen, die Sie installieren, beispielsweise 2.2.0 <http://spark.apache.org/docs/latest/configuration.html>.

Ebenso ist es möglich, Spark dynamisch bei jeder Anwendungsübermittlung zu konfigurieren. Eine Einstellung, mit der die Ressourcenzuordnung für einen Executor automatisch erhöht wird, ist über die Konfigurationsdatei `spark` verfügbar. Weitere Informationen finden Sie unter [Überschreiben der standardmäßigen Spark-Konfigurationseinstellungen](#).

## Ändern von Spark-Standardereinstellungen

Das folgende Beispiel zeigt, wie Sie einen Cluster mit der AWS CLI erstellen, bei dem `spark.executor.memory` auf 2G gesetzt ist.

### Note

Linux-Zeilensfortsetzungszeichen (`\`) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (`^`).

```
aws emr create-cluster --name "Spark cluster" --ami-version 3.11.0 \  
--applications Name=Spark, Args=[-d,spark.executor.memory=2G] --ec2-attributes \  
KeyName=myKey \  
--instance-type m3.xlarge --instance-count 3 --use-default-roles
```

## Aufträge an Spark übergeben

Verwenden Sie einen Schritt, um das Skript `spark-submit` auf Ihrem EMR-Cluster auszuführen und Arbeit an den Cluster zu senden. Fügen Sie den Schritt mit der folgenden Methode hinzu: `addJobFlowSteps` [AmazonElasticMapReduceClient](#)

```
AWSCredentials credentials = new BasicAWSCredentials(accessKey, secretKey);  
AmazonElasticMapReduceClient emr = new AmazonElasticMapReduceClient(credentials);  
StepFactory stepFactory = new StepFactory();  
AddJobFlowStepsRequest req = new AddJobFlowStepsRequest();
```

```
req.withJobFlowId("j-1K48XXXXXXHCB");

List<StepConfig> stepConfigs = new ArrayList<StepConfig>();

StepConfig sparkStep = new StepConfig()
    .withName("Spark Step")
    .withActionOnFailure("CONTINUE")
    .withHadoopJarStep(stepFactory.newScriptRunnerStep("/home/hadoop/spark/bin/spark-submit", "--class", "org.apache.spark.examples.SparkPi", "/home/hadoop/spark/lib/spark-examples-1.3.1-hadoop2.4.0.jar", "10"));

stepConfigs.add(sparkStep);
req.withSteps(stepConfigs);
AddJobFlowStepsResult result = emr.addJobFlowSteps(req);
```

## Überschreiben der standardmäßigen Spark-Konfigurationseinstellungen

Vielleicht möchten Sie die standardmäßigen Spark-Konfigurationswerte je nach Anwendung überschreiben. Sie können dies tun, wenn Sie Anwendungen mithilfe eines Schritts senden, wodurch im Wesentlichen Optionen an `spark-submit` übergeben werden. Sie möchten z. B. den Speicher ändern, der einem Executor-Prozess zugeordnet wurde. Dazu modifizieren Sie `spark.executor.memory`. Sie können den Schalter `--executor-memory` mit einem Argument wie dem folgenden angeben:

```
/home/hadoop/spark/bin/spark-submit --executor-memory 1g --class
org.apache.spark.examples.SparkPi /home/hadoop/spark/lib/spark-examples*.jar 10
```

Auf ähnliche Weise können Sie `--executor-cores` und `--driver-memory` optimieren. In einem Schritt würden Sie die folgenden Argumente bereitstellen:

```
--executor-memory 1g --class org.apache.spark.examples.SparkPi /home/hadoop/spark/lib/
spark-examples*.jar 10
```

Sie können mithilfe der Option `--conf` auch Einstellungen optimieren, die eventuell über keinen integrierten Schalter verfügen. Weitere Informationen zu anderen optimierbaren Einstellungen finden Sie im Thema [Dynamisches Laden von Spark-Eigenschaften](#) in der Apache-Spark-Dokumentation.



## Unterschiede zwischen dem DistCp S3-Hilfsprogramm und früheren AMI-Versionen von Amazon EMR

### In Amazon EMR unterstützte DistCp S3-Versionen

Die folgenden DistCp S3-Versionen werden in Amazon EMR AMI-Versionen unterstützt. DistCpS3-Versionen nach 1.0.7 befinden sich direkt auf den Clustern. Verwenden Sie die JAR-Datei in `/home/hadoop/lib`, um die neuesten Funktionen zu nutzen.

Version	Beschreibung	Datum der Veröffentlichung
1.0.8	Fügt die Optionen <code>--appendToLastFile</code> , <code>--requirePreviousManifest</code> und <code>--storageClass</code> hinzu.	3. Januar 2014
1.0.7	Fügt die Option <code>--s3ServerSideEncryption</code> hinzu.	2. Mai 2013
1.0.6	Fügt die Option <code>--s3Endpoint</code> hinzu.	6. August 2012
1.0.5	Verbessert die Fähigkeit, anzugeben, welche Version von S3 ausgeführt DistCp werden soll.	27. Juni 2012
1.0.4	Verbessert die Option <code>--deleteOnSuccess</code> .	19. Juni 2012
1.0.3	Bietet Support für die Optionen <code>--numberOfFiles</code> und <code>--startingIndex</code> .	12. Juni 2012
1.0.2	Verbessert die Dateibenennung bei der Verwendung von Gruppen.	6. Juni 2012
1.0.1	Erste Version von S3DistCp.	19. Januar 2012

### Fügen Sie einem Cluster einen DistCp S3-Kopierschritt hinzu

Um einem laufenden Cluster einen DistCp S3-Kopierschritt hinzuzufügen, geben Sie den folgenden Befehl ein, ersetzen Sie `j-3GYxxxxxx9I0K` durch Ihre Cluster-ID und ersetzen Sie `mybucket` durch Ihren Amazon S3 S3-Bucket-Namen.

**Note**

Linux-Zeilenfortsetzungszeichen (\) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (^).

```
aws emr add-steps --cluster-id j-3GYXXXXXX9I0K \
--steps Type=CUSTOM_JAR,Name="S3DistCp step",Jar=/home/hadoop/lib/emr-s3distcp-1.0.jar,
\
Args=["--s3Endpoint,s3-eu-west-1.amazonaws.com",\
"--src,s3://mybucket/logs/j-3GYXXXXXX9I0J/node/",\
"--dest,hdfs://output",\
"--srcPattern,.*[a-zA-Z,]+"]
```

**Example CloudFront Amazon-Protokolle in HDFS laden**

In diesem Beispiel werden CloudFront Amazon-Protokolle in HDFS geladen, indem einem laufenden Cluster ein Schritt hinzugefügt wird. Dabei wird das Komprimierungsformat von Gzip ( CloudFrontStandard) auf LZO geändert. Dies ist nützlich, da die mit LZO komprimierten Daten während der Dekomprimierung in mehrere Maps aufgeteilt werden können, sodass Sie nicht wie bei Gzip warten müssen, bis die Komprimierung abgeschlossen ist. Diese Vorgehensweise bietet eine bessere Leistung, wenn Sie die Daten mit Amazon EMR analysieren. Dieses Beispiel verbessert außerdem die Leistung, indem der in der Option --groupBy angegebene reguläre Ausdruck verwendet wird, um alle Protokolle für eine bestimmte Stunde in einer einzigen Datei zu kombinieren. Amazon-EMR-Cluster sind effizienter, wenn nur wenige große LZO-komprimierte Dateien verarbeitet werden anstelle von vielen kleinen mit Gzip komprimierten Dateien. Um LZO-Dateien aufzuteilen, müssen Sie sie indizieren und die Hadoop-izo-Drittanbieterbibliothek verwenden.

*Um CloudFront Amazon-Protokolle in HDFS zu laden, geben Sie den folgenden Befehl ein, ersetzen Sie j-3GYxxxxxx9I0K durch Ihre Cluster-ID und ersetzen Sie mybucket durch Ihren Amazon S3-Bucket-Namen.*

**Note**

Linux-Zeilenfortsetzungszeichen (\) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (^).

```
aws emr add-steps --cluster-id j-3GYXXXXXX9I0K \  
--steps Type=CUSTOM_JAR,Name="S3DistCp step",Jar=/home/hadoop/lib/emr-s3distcp-1.0.jar, \  
\   
Args=["--src,s3://mybucket/cf","--dest,hdfs:///local",\  
"--groupBy,. *XABCD12345678.([0-9]+-[0-9]+-[0-9]+-[0-9]+). *",\  
"--targetSize,128", \  
"--outputCodec,lzo","--deleteOnSuccess"]
```

Stellen Sie sich den Fall vor, in dem das vorherige Beispiel über die folgenden Protokolldateien ausgeführt wird. CloudFront

```
s3://DOC-EXAMPLE-BUCKET1/cf/XABCD12345678.2012-02-23-01.HLUS3JKx.gz  
s3://DOC-EXAMPLE-BUCKET1/cf/XABCD12345678.2012-02-23-01.I9CNAZrg.gz  
s3://DOC-EXAMPLE-BUCKET1/cf/XABCD12345678.2012-02-23-02.YRRwERSA.gz  
s3://DOC-EXAMPLE-BUCKET1/cf/XABCD12345678.2012-02-23-02.dshVLXFE.gz  
s3://DOC-EXAMPLE-BUCKET1/cf/XABCD12345678.2012-02-23-02.LpLfuShd.gz
```

S3 DistCp kopiert, verkettet und komprimiert die Dateien in die folgenden beiden Dateien, wobei der Dateiname durch die Übereinstimmung mit dem regulären Ausdruck bestimmt wird.

```
hdfs:///local/2012-02-23-01.lzo  
hdfs:///local/2012-02-23-02.lzo
```

# Neuerungen?

Auf dieser Seite werden die Änderungen und Funktionen beschrieben, die in den neuesten Versionen von Amazon EMR 7.x, 6.x und 5.x verfügbar sind.

Diese Versionshinweise sind auch auf den Seiten [Amazon EMR 7.1.0](#), [Amazon EMR 6.15.0](#) und [Amazon EMR 5.36.2](#) verfügbar, zusammen mit den Anwendungsversionen, Komponentenversionen und verfügbaren Konfigurationsklassifizierungen für jede Version.

- Versionshinweise zu früheren Versionen finden Sie unter [Amazon-EMR-Archiv mit Versionshinweisen](#).
- Wenn Sie Updates erhalten möchten, wenn eine neue Amazon-EMR-Version verfügbar ist, abonnieren Sie das [RSS-Feed für Amazon-EMR-Versionshinweise](#).

## Note

Spätere Versionen von Amazon EMR verwenden AWS Signature Version 4 (SigV4), um Anfragen an Amazon S3 zu authentifizieren. Wir empfehlen Ihnen, eine Amazon-EMR-Version zu verwenden, die SigV4 unterstützt, damit Sie auf neue S3-Buckets zugreifen und Unterbrechungen Ihrer Workloads vermeiden können. Weitere Informationen und eine Liste der Amazon-EMR-Versionen, die SigV4 unterstützen, finden Sie unter [Amazon EMR und AWS Signature Version 4](#).

## Amazon EMR 7.1.0 (neueste Version der 7.x-Serie)

Neue Amazon-EMR-Versionen werden über einen Zeitraum von mehreren Tagen in verschiedenen Regionen verfügbar gemacht, beginnend mit der ersten Region am ursprünglichen Veröffentlichungsdatum. Die neueste Version ist in diesem Zeitraum möglicherweise nicht in Ihrer Region verfügbar.

Die folgenden Versionshinweise enthalten Informationen für Amazon EMR Version 7.1.0. Die Änderungen beziehen sich auf 7.1.0.

## Neue Features

- Anwendungs-Upgrades – Amazon EMR 7.1.0 application upgrades include Livy 0.8.0, Trino 435, and ZooKeeper 3.9.1.
- [Austausch fehlerhafter Knoten](#) — In Amazon EMR 7.1.0 und höher ist der Austausch fehlerhafter Knoten standardmäßig aktiviert, sodass Amazon EMR Ihre fehlerhaften Knoten ordnungsgemäß ersetzt. Um zu vermeiden, dass Ihre bestehenden Workflows auf Amazon EMR-Versionen 7.0.0 und niedriger beeinträchtigt werden, wird der Austausch fehlerhafter Knoten deaktiviert, wenn Sie den Kündigungsschutz in Ihrem Cluster deaktiviert haben.
- [CloudWatch Agent](#) — Konfigurieren Sie den CloudWatch Agenten so, dass er zusätzliche Systemmetriken verwendet, Anwendungsmetriken hinzufügt und das Metrikziel mit der Amazon EMR-Konfigurations-API ändert.

## Bekannte Probleme

- Python 3.11 wird von EMR Studio nicht unterstützt.

## Änderungen, Verbesserungen und behobene Probleme

- Während Amazon EMR 7.1.0 Python 3.9 standardmäßig unterstützt, unterstützen Livy 0.8.0 und Spark in Amazon EMR 7.1.0 Python 3.11.
- Diese Version behebt das Problem, dass bei Verwendung PySpark mit Python Version 3.11 jede Zeile einzeln ausgeführt werden muss.
- Zeppelin-Upgrade — Amazon EMR 7.1.0 beinhaltet ein Upgrade von Zeppelin auf das SDK for Java v2. AWS Dieses Upgrade ermöglicht es einem Zeppelin S3-Notebook, einen Anbieter von kundenspezifischen Verschlüsselungsmaterialien zu akzeptieren. Das AWS SDK for Java v2 entfernt die EncryptionMaterialsProvider Schnittstelle. Wenn Sie ein Upgrade auf Amazon EMR 7.1.0 durchführen, müssen Sie die [Keyring-Schnittstelle](#) implementieren, wenn Sie eine benutzerdefinierte Verschlüsselung verwenden möchten. [Ein Beispiel für die Implementierung der Keyring-Schnittstelle finden Sie unter .java. KmsKeyring](#)
- Wenn Sie auf Amazon EMR Version 7.1.0 aktualisieren, ändern Sie Ihren benutzerdefinierten Schlüsselanbieter für die lokale Festplattenverschlüsselung, sodass Schlüssel mithilfe des AES-Algorithmus von AES/GCM/ generiert werden. NoPadding [Wenn Sie den Algorithmus nicht aktualisieren, schlägt die Clustererstellung möglicherweise mit der folgenden Fehlermeldung fehl. Weitere Informationen Local disk encryption failed on master instance](#)

[\(i-123456789\) due to internal error.](#) zum Erstellen eines benutzerdefinierten Schlüsselanbieters finden Sie unter [Erstellen eines benutzerdefinierten Schlüsselanbieters](#).

- Amazon EMR 7.1.0 verbessert die Ausfallsicherheit eines Knotens bei geringem Festplattenspeicher, indem die Logik zur Protokollkürzung für Dateien mit offenen Datei-Handles verbessert wird.
- Diese Version verbessert die Kodierungs- und Dekodierungslogik, um das Risiko einer Datenbeschädigung und eines Knotenausfalls zu minimieren, da Amazon EMR-Daemons beim Neustart eines Knotens Dateien lesen und schreiben.
- Wenn Sie einen Cluster mit der neuesten Patch-Version von Amazon EMR 5.36 oder höher, 6.6 oder höher oder 7.0 oder höher starten, verwendet Amazon EMR die neueste Version von Amazon Linux 2023 oder Amazon Linux 2 für das standardmäßige Amazon EMR-AMI. Weitere Informationen finden Sie unter [Verwenden des standardmäßigen Amazon-Linux-AMI für Amazon EMR](#).

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2023.3.2 240219.1	6.1.77-99.164.amzn2023	8. Mai 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), Kanada West (Calgary), AWS GovCloud (US-West), AWS GovCloud (US-Ost), China (Peking), China (Ningxia)

## Amazon EMR 6.15.0 (neueste Version der 6.x-Serie)

Neue Amazon-EMR-Versionen werden über einen Zeitraum von mehreren Tagen in verschiedenen Regionen verfügbar gemacht, beginnend mit der ersten Region am ursprünglichen Veröffentlichungsdatum. Die neueste Version ist in diesem Zeitraum möglicherweise nicht in Ihrer Region verfügbar.

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 6.15.0. Änderungen beziehen sich auf Version 6.14.0. Informationen zum Zeitplan für die Veröffentlichung finden Sie unter [6.15.0 Änderungsprotokoll](#).

## Neue Features

- Anwendungs-Upgrades – Amazon EMR 6.15.0 application upgrades include Apache Hadoop 3.3.6, Apache Hudi 0.14.0-amzn-0, Iceberg 1.4.0-amzn-0, and Trino 426.
- [Schnellere Starts für EMR-Cluster, die in EC2 ausgeführt werden](#) – Ein Amazon EMR kann jetzt bis zu 35 % schneller in einem EC2-Cluster gestartet werden. Mit dieser Verbesserung können die meisten Kunden ihre Cluster in 5 Minuten oder weniger starten.
- [CodeWhisperer für EMR Studio](#) — Sie können Amazon jetzt CodeWhisperer mit Amazon EMR Studio verwenden, um Empfehlungen in Echtzeit zu erhalten, während Sie Code einschreiben. JupyterLab CodeWhisperer kann Ihre Kommentare vervollständigen, einzelne Codezeilen fertigstellen, line-by-line Empfehlungen aussprechen und vollständig formatierte Funktionen generieren.
- [Schnellere Auftrag-Neustartzeiten mit Flink](#) – Mit Amazon EMR 6.15.0 und höher sind mehrere neue Mechanismen für Apache Flink verfügbar, um die Auftrag-Neustartzeit bei Aufgabenwiederherstellungs- oder -skalierungsvorgängen zu verkürzen. Dadurch wird die Geschwindigkeit der Wiederherstellung und des Neustarts von Ausführungsdiagrammen optimiert, um die Auftragsstabilität zu verbessern.
- Detaillierte [Zugriffskontrolle auf Tabellenebene für Open-Table-Formate](#) — Wenn Sie mit Amazon EMR 6.15.0 und höher Spark-Jobs auf Amazon EMR auf EC2-Clustern ausführen, die auf Daten im AWS Glue-Datenkatalog zugreifen, können Sie damit Berechtigungen auf Tabellen-, Zeilen-, Spalten- und Zellenebene auf Hudi-, Iceberg- oder Delta Lake-basierte Tabellen anwenden. AWS Lake Formation
- Hadoop-Upgrade – Amazon EMR 6.15.0 beinhaltet ein Upgrade von Apache Hadoop auf Version 3.3.6. Hadoop 3.3.6 war die neueste Version zum Zeitpunkt der Bereitstellung von Amazon EMR 6.15, die im Juni 2023 von Apache veröffentlicht wurde. Frühere Versionen von Amazon EMR (6.9.0 bis 6.14.x) verwendeten Hadoop 3.3.3.

Das Upgrade umfasst Hunderte Verbesserungen und Fehlerbehebungen sowie Features wie rekonfigurierbare Datenknotenparameter, die DFSAdmin-Option zum Initiieren von Massenrekonfigurationsvorgängen für alle Live-Datenknoten und eine vektorbasierte API, die es suchlastigen Lesern ermöglicht, mehrere Lesebereiche anzugeben. Hadoop 3.3.6 bietet außerdem Unterstützung für HDFS-APIs und Semantik für sein Write Ahead Log (WAL), sodass HBase in anderen Speichersystemimplementierungen ausgeführt werden kann. Weitere Informationen finden Sie in den Änderungsprotokollen für Versionen [3.3.4](#), [3.3.5](#) und [3.3.6](#) in der Apache-Hadoop-Dokumentation.



- Support für AWS SDK for Java, Version 2 — Amazon EMR 6.15.0-Anwendungen können AWS SDK for Java Java-Versionen [1.12.569](#) oder [2.20.160](#) verwenden, wenn die Anwendung Version 2 unterstützt. Das AWS SDK for Java 2.x ist eine umfassende Neufassung der Codebasis von Version 1.x. Sie basiert auf Java 8+ und fügt mehrere häufig angeforderte Funktionen hinzu. Dies umfasst die Unterstützung von blockierungsfreiem E/A und die Möglichkeit, zur Laufzeit eine andere HTTP-Implementierung einzubinden. Weitere Informationen, einschließlich eines Migrationshandbuchs von SDK für Java v1 zu v2, finden Sie im Handbuch [AWS SDK für Java, Version 2](#).

## Änderungen, Verbesserungen und behobene Probleme

- Um Ihre EMR-Cluster mit hoher Verfügbarkeit zu verbessern, ermöglicht diese Version Konnektivität zu Amazon-EMR-Daemons auf lokalen Hosts, die IPv6-Endpunkte verwenden.
- Diese Version aktiviert TLS 1.2 für die Kommunikation mit ZooKeeper Provisioned auf allen primären Knoten Ihres Hochverfügbarkeitsclusters.
- Diese Version verbessert die Verwaltung von ZooKeeper Transaktionsprotokolldateien, die auf primären Knoten verwaltet werden, um Szenarien zu minimieren, in denen die Protokolldateien über die Grenzen hinaus wachsen und den Clusterbetrieb unterbrechen.
- Diese Version verbessert die Widerstandsfähigkeit der knoteninternen Kommunikation von EMR-Clustern mit hoher Verfügbarkeit. Durch diese Verbesserung wird die Wahrscheinlichkeit, dass Bootstrap-Aktionen oder Cluster-Startfehler auftreten, verringert.
- Tez in Amazon EMR 6.15.0 führt Konfigurationen ein, die Sie angeben können, um die Eingabe-Splits in einem gruppierten Tez-Split asynchron zu öffnen. Dies führt zu einer schnelleren Leistung von Leseabfragen, wenn ein einzelner gruppierter Tez-Split eine große Anzahl von Eingabe-Splits enthält. Weitere Informationen finden Sie unter [Asynchrone Öffnung von Tez-Splits](#).
- Wenn Sie einen Cluster mit der neuesten Patch-Version von Amazon EMR 5.36 oder höher, 6.6 oder höher oder 7.0 oder höher starten, verwendet Amazon EMR die neueste Version von Amazon Linux 2023 oder Amazon Linux 2 für das standardmäßige Amazon EMR-AMI. Weitere Informationen finden Sie unter [Verwenden des standardmäßigen Amazon-Linux-AMI für Amazon EMR](#).

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 223.0	4,14.336	8. März 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			AWS GovCloud (US-Ost), China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024131.0	4,14.336	14. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 124.0	4,14.336	7. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024109.0	4,14.334	24. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 218.0	4,14.330	2. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsReleaseLabel (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202306.0	4,14.330	22. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 116.0	4,14.328	11. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023101.0	4,14.327	13. November 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),



OsReleaseLabel (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

## Amazon EMR 5.36.2 (neueste Version der 5.x-Serie)

Neue Amazon-EMR-Versionen werden über einen Zeitraum von mehreren Tagen in verschiedenen Regionen verfügbar gemacht, beginnend mit der ersten Region am ursprünglichen Veröffentlichungsdatum. Die neueste Version ist in diesem Zeitraum möglicherweise nicht in Ihrer Region verfügbar.

Die folgenden Versionshinweise enthalten Informationen für Amazon EMR Version 5.36.2. Die Änderungen beziehen sich auf 5.36.1. Informationen zum Zeitplan der Veröffentlichung finden Sie im [Änderungsprotokoll](#).

### Änderungen, Verbesserungen und behobene Probleme

- Diese Version verbessert die Cluster-Scale-Down-Logik, sodass Amazon EMR die Kernknoten nicht unter die HDFS-Replikationsfaktor-Einstellung für den Cluster herunterskaliert. Diese Verbesserung erfüllt die Anforderungen an die Datenredundanz und verringert die Wahrscheinlichkeit, dass ein Skalierungsvorgang zum Stillstand kommt.
- Diese Version fügt dem Cluster-Skalierungs-Workflow für die Ausführung Presto oder Trino einen neuen Wiederholungsmechanismus hinzu. Diese Verbesserung reduziert das Risiko, dass die Clustergrößenänderung aufgrund eines einzigen fehlgeschlagenen Größenänderungsvorgangs auf unbestimmte Zeit ausgeführt wird. Sie verbessert auch die Clusterauslastung, da Ihr Cluster schneller hoch- und herunterskaliert wird.
- Behebt ein Problem, bei dem Cluster-Scale-Down-Operationen möglicherweise zum Stillstand kommen, während Amazon EMR einen Core-Knoten ordnungsgemäß außer Betrieb nimmt und dieser vor der vollständigen Außerbetriebnahme fehlerhaft wird.
- Verbessert die Stabilität eines Knotens in einem Hochverfügbarkeitscluster mit mehreren Primärknoten, wenn Amazon EMR einen einzelnen Knoten neu startet.

- Optimiert die Protokollverwaltung mit Amazon EMR, das auf Amazon EC2 ausgeführt wird. Infolgedessen können Sie eine leichte Senkung der Speicherkosten für Ihre Cluster-Protokolle feststellen.
- Verbessert die Verwaltung von ZooKeeper Transaktionsprotokolldateien, die auf primären Knoten verwaltet werden, um Szenarien zu minimieren, in denen die Protokolldateien über die Grenzen hinaus wachsen und Clusteroperationen unterbrechen.
- Behebt einen seltenen Fehler, der dazu führen kann, dass ein Hochverfügbarkeitscluster mit mehreren Primärknoten ausfällt, weil er nicht mit dem Yarn kommunizieren kann.  
ResourceManager
- Wenn Sie einen Cluster mit der neuesten Patch-Version von Amazon EMR 5.36 oder höher, 6.6 oder höher oder 7.0 oder höher starten, verwendet Amazon EMR die neueste Version von Amazon Linux 2023 oder Amazon Linux 2 für das standardmäßige Amazon EMR-AMI. Weitere Informationen finden Sie unter [Verwenden des standardmäßigen Amazon-Linux-AMI für Amazon EMR](#).

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 503.0	4,14.343	xxxxxx, 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West), (US-Ost), China (Peking), China (Ningxia)

## Amazon EMR und AWS Signature Version 4

Amazon EMR-Versionen verwenden AWS Signature Version 4 (SigV4) zur Authentifizierung von Anfragen an Amazon S3. Buckets, die nach dem 24. Juni 2020 in Amazon S3 erstellt wurden, unterstützen keine Anfragen, die mit Signature Version 2 (SIGv2) signiert wurden. Buckets, die am oder vor dem 24. Juni 2020 erstellt wurden, unterstützen weiterhin SIGv2. Wir empfehlen, dass Sie zu einer Amazon-EMR-Version migrieren, die SigV4 unterstützt. So können Sie weiterhin auf neue S3-Buckets zugreifen und mögliche Unterbrechungen Ihrer Workloads vermeiden.

Wenn Sie Anwendungen verwenden, die in Amazon EMR enthalten sind, wie Apache Spark, Apache Hive und Presto, müssen Sie Ihren Anwendungscode nicht ändern, um SigV4 zu verwenden. Wenn Sie benutzerdefinierte Anwendungen verwenden, die nicht in Amazon EMR enthalten sind, müssen Sie möglicherweise Ihren Code aktualisieren, um SigV4 zu verwenden. Weitere

Informationen finden Sie unter [Umstieg von Signature Version 2 auf Signature Version 4](#) im Amazon-S3-Benutzerhandbuch.

Die folgenden Amazon-EMR-Versionen unterstützen SigV4: emr-4.7.4, emr-4.8.5, emr-4.9.6, emr-4.10.1, emr-5.1.1, emr-5.2.3, emr-5.3.2, emr-5.4.1, emr-5.5.4, emr-5.6.1, emr-5.7.1, emr-5.8.3, emr-5.9.1, emr-5.10.1, emr-5.11.4, emr-5.12.3, emr-5.13.1, emr-5.14.2, emr-5.15.1, emr-5.16.1, emr-5.17.2, emr-5.18.1, emr-5.19.1, emr-5.20.1, emr-5.21.2, and emr-5.22.0 and higher. Alle 6.x- und 7.x-Versionen unterstützen SigV4.

## Ansatz zur Minderung von CVE-2021-44228

### Note

Für Amazon-EMR-Version 6.9.0 und höher verwenden alle von Amazon EMR installierten Komponenten, die Log4j-Bibliotheken verwenden, Log4j Version 2.17.1 oder höher.

### Amazon EMR läuft auf EC2

Das in [CVE-2021-44228](#) behandelte Problem ist für Apache-Log4j-Kernversionen zwischen 2.0.0 und 2.14.1 relevant, wenn Eingaben aus nicht vertrauenswürdigen Quellen verarbeitet werden. Amazon-EMR-Cluster, die mit Amazon-EMR-5.x-Versionen bis 5.34.0 und EMR-6.x-Versionen bis Amazon EMR 6.5.0 gestartet wurden, enthalten Open-Source-Frameworks wie Apache Hive, Flink, HUDI, Presto und Trino, die diese Versionen von Apache Log4j verwenden. Viele Kunden verwenden jedoch die auf ihren Amazon-EMR-Clustern installierten Open-Source-Frameworks, um Eingaben aus nicht vertrauenswürdigen Quellen zu verarbeiten und zu protokollieren.

Wir empfehlen, dass Sie die „Lösung Amazon EMR-Bootstrap-Aktion für Log4j CVE-2021-44228“ wie im folgenden Abschnitt beschrieben anwenden. Diese Lösung befasst sich auch mit CVE-2021-45046.

### Note

Die Bootstrap-Aktionsskripte für Amazon EMR wurden am 7. September 2022 aktualisiert und enthalten nun schrittweise Bugfixes und Verbesserungen für Oozie. Wenn Sie Oozie verwenden, sollten Sie die aktualisierte Amazon-EMR-Bootstrap-Aktionslösung anwenden, die im folgenden Abschnitt beschrieben wird.

## Amazon EMR in EKS

Wenn Sie [Amazon EMR in EKS](#) mit Standardkonfiguration verwenden, sind Sie von dem in CVE-2021-44228 beschriebenen Problem nicht betroffen und Sie müssen die im Abschnitt [Amazon-EMR-Bootstrap-Aktionslösung für Log4J CVE-2021-44228 und CVE-2021-45046](#) beschriebene Lösung nicht anwenden. Für Amazon EMR auf EKS verwendet die Amazon-EMR-Laufzeit für Spark Apache Log4j Version 1.2.17. Wenn Sie Amazon EMR in EKS verwenden, sollten Sie die Standardeinstellung für die `log4j.appendix` Komponente nicht auf `log` ändern.

## Amazon-EMR-Bootstrap-Aktionslösung für Log4J CVE-2021-44228 und CVE-2021-45046

Diese Lösung bietet eine Amazon-EMR-Bootstrap-Aktion, die auf Ihre Amazon-EMR-Cluster angewendet werden muss. Für jede Amazon-EMR-Version finden Sie unten einen Link zu einem Bootstrap-Aktionsskript. Sie sollten die folgenden Schritte ausführen, um diese Bootstrap-Aktion anzuwenden:

1. Kopieren Sie das Skript, das Ihrer Amazon-EMR-Version entspricht, in einen lokalen S3-Bucket in Ihrem AWS-Konto. Bitte stellen Sie sicher, dass Sie ein Bootstrap-Skript verwenden, das für Ihre Amazon-EMR-Version spezifisch ist.
2. Richten Sie eine Bootstrap-Aktion für Ihre EMR-Cluster ein, um das in Ihren S3-Bucket kopierte Skript gemäß den in der [EMR-Dokumentation](#) beschriebenen Anweisungen auszuführen. Wenn Sie andere Bootstrap-Aktionen für Ihre EMR-Cluster konfiguriert haben, stellen Sie sicher, dass dieses Skript als erstes Bootstrap-Aktionsskript eingerichtet ist, das ausgeführt wird.
3. Beenden Sie bestehende EMR-Cluster und starten Sie neue Cluster mit dem Bootstrap-Aktionsskript. AWS empfiehlt, dass Sie die Bootstrap-Skripte in Ihrer Testumgebung testen und Ihre Anwendungen validieren, bevor Sie sie in Ihrer Produktionsumgebung anwenden. Wenn Sie nicht die neueste Version für eine EMR-Nebenversion (z. B. 6.3.0) verwenden, müssen Sie die neueste Version (z. B. 6.3.1) verwenden und dann die oben beschriebene Lösung anwenden.

### CVE-2021-44228 & CVE-2021-45046 – Bootstrap-Skripte für Amazon-EMR-Versionen

Amazon-EMR-Version nummer	Skript-Speicherort	Datum der Veröffentlichung des Skripts
6.5.0	<code>s3://elasticmapreduce/ bootstrap-actions/</code>	24. März 2022

Amazon-EMR-Version nummer	Skript-Speicherort	Datum der Veröffentlichung des Skripts
	<code>log4j/patch-log4j-emr-6.5.0-v2.sh</code>	
6.4.0	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-6.4.0-v2.sh</code>	24. März 2022
6.3.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-6.3.1-v2.sh</code>	24. März 2022
6.2.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-6.2.1-v2.sh</code>	24. März 2022
6.1.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-6.1.1-v2.sh</code>	14. Dezember 2021
6.0.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-6.0.1-v2.sh</code>	14. Dezember 2021
5,34,0	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.34.0-v2.sh</code>	12. Dezember 2021

Amazon-EMR-Version nummer	Skript-Speicherort	Datum der Veröffentlichung des Skripts
5,33,1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.33.1-v2.sh</code>	12. Dezember 2021
5.32,1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.32.1-v2.sh</code>	13. Dezember 2021
5.31,1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.31.1-v2.sh</code>	13. Dezember 2021
5,30,2	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.30.2-v2.sh</code>	14. Dezember 2021
5.29,0	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.29.0-v2.sh</code>	14. Dezember 2021
5,28,1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.28.1-v2.sh</code>	15. Dezember 2021

Amazon-EMR-Version nummer	Skript-Speicherort	Datum der Veröffentlichung des Skripts
5,27,1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.27.1-v2.sh</code>	15. Dezember 2021
5,26,0	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.26.0-v2.sh</code>	15. Dezember 2021
5,25,0	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.25.0-v2.sh</code>	15. Dezember 2021
5,24,1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.24.1-v2.sh</code>	15. Dezember 2021
5.23,1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.23.1-v2.sh</code>	15. Dezember 2021
5,22,0	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.22.0-v2.sh</code>	15. Dezember 2021



Amazon-EMR-Version nummer	Skript-Speicherort	Datum der Veröffentlichung des Skripts
5,21,2	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.21.2-v2.sh</code>	15. Dezember 2021
5,20,1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.20.1-v2.sh</code>	15. Dezember 2021
5.19,1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.19.1-v2.sh</code>	15. Dezember 2021
5.18.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.18.1-v2.sh</code>	15. Dezember 2021
5.17.2	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.17.2-v2.sh</code>	15. Dezember 2021
5.16.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.16.1-v2.sh</code>	15. Dezember 2021

Amazon-EMR-Version nummer	Skript-Speicherort	Datum der Veröffentlichung des Skripts
5.15.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.15.1-v2.sh</code>	15. Dezember 2021
5.14.2	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.14.2-v2.sh</code>	15. Dezember 2021
5.13.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.13.1-v2.sh</code>	15. Dezember 2021
5.12.3	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.12.3-v2.sh</code>	15. Dezember 2021
5.11.4	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.11.4-v2.sh</code>	15. Dezember 2021
5.10.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.10.1-v2.sh</code>	15. Dezember 2021

Amazon-EMR-Version nummer	Skript-Speicherort	Datum der Veröffentlichung des Skripts
5.9.1	s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.9.1-v2.sh	15. Dezember 2021
5.8.3	s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.8.3-v2.sh	15. Dezember 2021
5.7.1	s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.7.1-v2.sh	15. Dezember 2021

EMR-Version	Letzte Version vom Dezember 2021
6.3.0	6.3.1
6.2.0	6.2.1
6.1.0	6.1.1
6.0.0	6.0.1
5.33,0	5.33,1
5,32,0	5,32,1
5,31,0	5,31,1
5.30.0 oder 5.30.1	5,30,2
5,28,0	5,28,1

EMR-Version	Letzte Version vom Dezember 2021
5,27,0	5,27,1
5,24,0	5,24,1
5,23,0	5.23,1
5.21.0 oder 5.21.1	5.21.2
5,20,0	5,20,1
5.19,0	5.19.1
5.18.0	5.18.1
5.17.0 oder 5.17.1	5.17.2
5.16.0	5.16.1
5.15.0	5.15,1
5.14.0 oder 5.14.1	5.14.2
5.13.0	5.13.1
5.12,0, 5.12,1, 5.12,2	5.12.3
5.11.0, 5.11.1, 5.11.2, 5.11.3	5.11.4
5.9.0	5.9.1
5.8.0, 5.8.1, 5.8.2	5.8.3
5.7.0	5.7.1

## Häufig gestellte Fragen

- Sind EMR-Versionen, die älter als EMR 5 sind, von CVE-2021-44228 betroffen?

Nein. EMR-Versionen vor EMR-Version 5 verwenden Log4J-Versionen, die älter als 2.0 sind.

- Befasst sich diese Lösung mit CVE-2021-45046?

Ja, diese Lösung adressiert auch CVE-2021-45046.

- Unterstützt die Lösung benutzerdefinierte Anwendungen, die ich auf meinen EMR-Clustern installiere?

Das Bootstrap-Skript aktualisiert nur JAR-Dateien, die von EMR installiert wurden. Wenn Sie benutzerdefinierte Anwendungen und JAR-Dateien auf Ihren EMR-Clustern über Bootstrap-Aktionen, als an Ihre Cluster übermittelte Schritte, mithilfe eines benutzerdefinierten Amazon-Linux-AMI oder über einen anderen Mechanismus installieren und ausführen, wenden Sie sich bitte an Ihren Anwendungsanbieter, um festzustellen, ob Ihre benutzerdefinierten Anwendungen von CVE-2021-44228 betroffen sind, und suchen Sie nach einer geeigneten Lösung.

- Wie sollte ich mit [benutzerdefinierten Docker-Images](#) mit EMR in EKS umgehen?

Wenn Sie benutzerdefinierte Anwendungen mithilfe [benutzerdefinierter Docker-Images](#) zu Amazon EMR in EKS hinzufügen oder Jobs mit benutzerdefinierten Anwendungsdateien an Amazon EMR in EKS senden, ermitteln Sie gemeinsam mit dem Anwendungsanbieter, ob Ihre benutzerdefinierten Anwendungen von CVE-2021-44228 betroffen sind, und ermitteln Sie eine geeignete Lösung.

- Wie funktioniert das Bootstrap-Skript, um das in CVE-2021-44228 und CVE-2021-45046 beschriebene Problem zu entschärfen?

Das Bootstrap-Skript aktualisiert die EMR-Startup-Anweisungen, indem es einen neuen Befehlssatz hinzufügt. Diese neuen Anweisungen löschen die JndiLookup Klassendateien, die über Log4j von allen von EMR installierten Open-Source-Frameworks verwendet werden. Dies folgt der [von Apache veröffentlichten Empfehlung](#) zur Behebung der Log4j-Probleme.

- Gibt es ein Update für EMR, das Log4j-Versionen 2.17.1 oder höher verwendet?

EMR-5-Versionen bis Version 5.34 und EMR-6-Versionen bis Version 6.5 verwenden ältere Versionen von Open-Source-Frameworks, die mit den neuesten Versionen von Log4j nicht kompatibel sind. Wenn Sie diese Versionen weiterhin verwenden, empfehlen wir Ihnen, die Bootstrap-Aktion anzuwenden, um die in den CVEs erörterten Probleme zu beheben. Nach EMR-5-Version 5.34 und EMR-6-Version 6.5 werden Anwendungen, die Log4j 1.x und Log4j 2.x verwenden, auf Log4j 1.2.17 (oder höher) bzw. Log4j 2.17.1 (oder höher) aktualisiert, sodass die oben genannten Bootstrap-Aktionen nicht erforderlich sind, um die CVE-Probleme zu beheben.

- Sind EMR-Versionen von CVE-2021-45105 betroffen?

Die von Amazon EMR mit den Standardkonfigurationen von EMR installierten Anwendungen sind von CVE-2021-45105 nicht betroffen. Unter den von Amazon EMR installierten Anwendungen verwendet nur Apache Hive Apache Log4j mit [Kontext-Lookups](#) und verwendet kein vom Standard abweichendes Muster-Layout, sodass unangemessene Eingabedaten verarbeitet werden können.

- Ist Amazon EMR von einer der folgenden CVE-Offenlegungen betroffen?

Die folgende Tabelle enthält eine Liste von CVEs, die sich auf Log4j beziehen, und gibt an, ob sich jede CVE auf Amazon EMR auswirkt. Die Informationen in dieser Tabelle gelten nur, wenn Anwendungen von Amazon EMR mit den Standardkonfigurationen installiert werden.

CVE	Auswirkungen EMR	Hinweise
CVE-2022-23302	Nein	Amazon EMR richtet Log4j JMSSink nicht ein
CVE-2022-23305	Nein	Amazon EMR richtet Log4j JdbcAppender nicht ein
CVE-2022-23307	Nein	Amazon EMR richtet Log4j Chainsaw nicht ein
CVE-2020-9493	Nein	Amazon EMR richtet Log4j Chainsaw nicht ein
CVE-2021-44832	Nein	Amazon EMR richtet Log4j JdbcAppender nicht mit einer JNDI-Verbindungszeichenfolge ein
CVE-2021-4104	Nein	Amazon EMR verwendet Log4j JmsAppender nicht
CVE-2020-9488	Nein	Die von Amazon EMR installierten Anwendungen verwenden Log4j SMTPAppender nicht

CVE	Auswirkungen EMR	Hinweise
CVE-2019-17571	Nein	Amazon EMR blockiert den öffentlichen Zugriff auf Cluster und startet nicht SocketServer
CVE-2019-17531	Nein	Es wird empfohlen, ein Upgrade auf die neueste Amazon-EMR-Release-Version durchzuführen. Amazon EMR 5.33.0 und höher verwenden jackson-databind 2.6.7.4 oder höher, und EMR 6.1.0 und höher verwenden jackson-databind 2.10.0 oder höher. Diese Versionen von jackson-databind sind vom CVE nicht betroffen.

## Amazon-EMR-Archiv mit Versionshinweisen

Versionshinweise für alle Amazon-EMR-Versionen sind unten verfügbar. Umfassende Versionsinformationen für jede Version finden Sie unter [Amazon-EMR-6.x-Versionen](#), [Amazon-EMR-5.x-Versionen](#) und [Amazon-EMR-4.x-Versionen](#).

Wenn Sie Updates erhalten möchten, wenn eine neue Amazon-EMR-Version verfügbar ist, abonnieren Sie das [RSS-Feed für Amazon-EMR-Versionshinweise](#).

### Version 6.14.0

Die folgenden Versionshinweise enthalten Informationen über Amazon EMR Version 6.14.0. Änderungen beziehen sich auf Version 6.13.0. Informationen zum Zeitplan für die Veröffentlichung finden Sie unter [6.14.0 Änderungsprotokoll](#).

## Neue Features

- Amazon EMR 6.14.0 supports Apache Spark 3.4.1, Apache Spark RAPIDS 23.06.0-amzn-2, Flink 1.17.1, Iceberg 1.3.1, and Trino 422.
- [Amazon EMR Managed Scaling](#) ist jetzt in der Region ap-southeast-3 Asien-Pazifik (Jakarta) für Cluster verfügbar, die Sie mit Amazon EMR 6.14.0 und höher erstellen.

## Änderungen, Verbesserungen und behobene Probleme

- Die Version 6.14.0 optimiert die Protokollverwaltung mit Amazon EMR, das auf Amazon EC2 ausgeführt wird. Infolgedessen können Sie eine leichte Senkung der Speicherkosten für Ihre Cluster-Protokolle feststellen.
- Die Version 6.14.0 verbessert den Skalierungs-Workflow, um verschiedene Core-Instances zu berücksichtigen, deren Größe für ihre Amazon-EBS-Volumes erheblich variiert. Diese Verbesserung gilt nur für Core-Knoten; Herunterskalierungs-Operationen für Aufgabenknoten sind davon nicht betroffen.
- Die Version 6.14.0 verbessert die Interaktion von Amazon EMR mit Open-Source-Anwendungen wie Apache Hadoop YARN ResourceManager and HDFS NameNode. Diese Verbesserung reduziert das Risiko von Betriebsverzögerungen bei der Cluster-Skalierung und verringert Startup-Fehler, die aufgrund von Verbindungsproblemen mit den Open-Source-Anwendungen auftreten.
- Die Version 6.14.0 optimiert die Anwendungsinstallation beim Clusterstart. Dies verbessert die Cluster-Startup-Zeiten für bestimmte Kombinationen von Amazon-EMR-Anwendungen.
- Die Version 6.14.0 behebt ein Problem, bei dem Cluster-Herunterskalierungs-Operationen möglicherweise zum Stillstand kommen, wenn ein Cluster, der in einer VPC mit einer benutzerdefinierten Domain ausgeführt wird, auf einen Core- oder Aufgabenknoten-Neustart stößt.
- Wenn Sie einen Cluster mit der neuesten Patch-Version von Amazon EMR 5.36 oder höher, 6.6 oder höher oder 7.0 oder höher starten, verwendet Amazon EMR die neueste Version von Amazon Linux 2023 oder Amazon Linux 2 für das standardmäßige Amazon EMR-AMI. Weitere Informationen finden Sie unter [Verwenden des standardmäßigen Amazon-Linux-AMI für Amazon EMR](#).



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 223.0	4,14.336	8. März 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			AWS GovCloud (US-Ost), China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024131.0	4,14.336	14. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 124.0	4,14.336	7. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024109.0	4,14.334	24. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 218.0	4,14.330	2. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsReleaseLabel (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202306.0	4,14.330	22. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 116.0	4,14.328	11. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsReleaseLabel (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023101.0	4,14.327	17. November 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 4,14.322 906.0		11. September 2023	China (Peking), China (Ningxia)  USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv)



## Version 6.13.0

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 6.13.0. Änderungen beziehen sich auf Version 6.12.0. Informationen zum Zeitplan für die Veröffentlichung finden Sie unter [6.13.0 Änderungsprotokoll](#).

### Neue Features

- Amazon EMR 6.13.0 supports Apache Spark 3.4.1, Apache Spark RAPIDS 23.06.0-amzn-1, CUDA Toolkit 11.8.0, and JupyterHub 1.5.0.

### Änderungen, Verbesserungen und behobene Probleme

- Die Version 6.13.0 verbessert den Amazon-EMR-Protokollverwaltungs-Daemon, um sicherzustellen, dass alle Protokolle in regelmäßigen Abständen auf Amazon S3 hochgeladen werden, wenn ein Befehl zur Clusterbeendigung ausgegeben wird. Dies ermöglicht schnellere Clusterbeendigungen.
- Die Version 6.13.0 verbessert die Protokollverwaltungsfunktionen von Amazon EMR, um einen konsistenten und zeitnahen Upload aller Protokolldateien auf Amazon S3 sicherzustellen. Davon profitieren vor allem langlaufende EMR-Cluster.
- Wenn Sie einen Cluster mit der neuesten Patch-Version von Amazon EMR 5.36 oder höher, 6.6 oder höher oder 7.0 oder höher starten, verwendet Amazon EMR die neueste Version von Amazon Linux 2023 oder Amazon Linux 2 für das standardmäßige Amazon EMR-AMI. Weitere Informationen finden Sie unter [Verwenden des standardmäßigen Amazon-Linux-AMI für Amazon EMR](#).

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 223.0	4,14.336	8. März 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			<p>Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost), China (Peking), China (Ningxia), Kanada West (Calgary)</p>

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024131.0	4,14.336	14. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 124.0	4,14.336	7. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024109.0	4,14.334	24. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 218.0	4,14.330	2. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsReleaseLabel (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202306.0	4,14.330	22. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 116.0	4,14.328	11. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsReleaseLabel (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023101.0	4,14.327	16. November 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023020.1	4,14.326	7. November 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023012.1	4,14.326	26. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 926.0	4,14.322	19. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)
2.0.2023 906.0	4,14.322	04. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 808,0	4,14.320	24. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv)

## Version 6.12.0

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 6.12.0. Änderungen beziehen sich auf Version 6.11.0. Informationen zum Zeitplan für die Veröffentlichung finden Sie unter [6.12.0 Änderungsprotokoll](#).

### Neue Features

- Amazon EMR 6.12.0 unterstützt Apache Spark 3.4.0, Apache Spark RAPIDS 23.06.0-amzn-0, CUDA 11.8.0, Apache Hudi 0.13.1-amzn-0, Apache Iceberg 1.3.0-amzn-0, Trino 414, und PrestoDB 0.281.
- Amazon EMR-Versionen 6.12.0 und höher unterstützen die LDAP-Integration mit Apache Livy, Apache Hive bis HiveServer 2 (HS2), Trino, Presto und Hue. Sie können Apache Spark und Apache Hadoop auch auf einem EMR-Cluster installieren, der 6.12.0 oder höher verwendet, und sie für die Verwendung von LDAP konfigurieren. Weitere Informationen finden Sie unter [Verwenden von Active-Directory- oder LDAP-Servern für die Authentifizierung mit Amazon EMR](#).

### Änderungen, Verbesserungen und behobene Probleme

- Amazon-EMR-Versionen 6.12.0 und höher bieten Java-11-Laufzeitunterstützung für Flink. Weitere Informationen finden Sie unter [Flink für die Ausführung mit Java 11 konfigurieren](#).
- Die Version 6.12.0 fügt dem Cluster-Skalierungs-Workflow für EMR-Cluster, auf denen Presto oder Trino ausgeführt werden, einen neuen Wiederholungsmechanismus hinzu. Diese Verbesserung verringert das Risiko, dass die Clustergrößenänderung aufgrund eines einzigen fehlgeschlagenen Größenänderungsvorgangs auf unbestimmte Zeit zum Stillstand kommt. Sie verbessert auch die Clusterauslastung, da Ihr Cluster schneller hoch- und herunterskaliert wird.
- Die Version 6.12.0 behebt ein Problem, bei dem Cluster-Herunterskalierungs-Operationen zum Stillstand kommen können, wenn ein Core-Knoten, der einer ordnungsgemäßen Außerbetriebnahme unterzogen wird, aus irgendeinem Grund nicht mehr funktionsfähig ist, bevor er vollständig außer Betrieb genommen wird.
- Die Version 6.12.0 verbessert die Cluster-Herunterskalierungs-Logik, sodass Ihr Cluster nicht versucht, die Core-Knoten unter die für den Cluster eingestellte HDFS-Replikationsfaktor-Einstellung herunterzuskalieren. Dies entspricht Ihren Anforderungen an die Datenredundanz und verringert die Wahrscheinlichkeit, dass ein Skalierungsvorgang zum Stillstand kommt.
- Die Version 6.12.0 verbessert die Leistung und Effizienz des Health Monitoring Service für Amazon EMR, indem die Geschwindigkeit erhöht wird, mit der Statusänderungen für Instances protokolliert



werden. Diese Verbesserung verringert das Risiko einer Leistungseinbuße bei Cluster-Knoten, auf denen mehrere benutzerdefinierte Client-Tools oder Drittanbieteranwendungen ausgeführt werden.

- Die Version 6.12.0 verbessert die Leistung des On-Cluster-Protokollverwaltungs-Daemons für Amazon EMR. Daher ist die Wahrscheinlichkeit einer Leistungsverschlechterung bei EMR-Clustern, die Schritte mit hoher Parallelität ausführen, geringer.
- Mit der Amazon-EMR-Version 6.12.0 wurde der Protokoll-Management-Daemon aktualisiert, um alle Protokolle, die aktiv verwendet werden, mit offenen Datei-Handles auf dem lokalen Instance-Speicher und die zugehörigen Prozesse zu identifizieren. Dieses Upgrade stellt sicher, dass Amazon EMR die Dateien ordnungsgemäß löscht und Speicherplatz zurückgewinnt, nachdem die Protokolle in Amazon S3 archiviert wurden.
- Die Version 6.12.0 beinhaltet eine Erweiterung des Protokollverwaltungs-Daemons, die leere, unbenutzte Schrittverzeichnisse im lokalen Cluster-Dateisystem löscht. Eine zu große Anzahl leerer Verzeichnisse kann die Leistung der Amazon-EMR-Daemons beeinträchtigen und zu einer Überauslastung der Festplatte führen.
- Die Version 6.12.0 ermöglicht die Protokollrotation für YARN Timeline Server-Protokolle. Dadurch werden Szenarien mit übermäßiger Festplattenauslastung minimiert, insbesondere bei Clustern mit langer Laufzeit.
- Die Standardgröße des Root-Volumes wurde in Amazon EMR 6.10.0 und höher auf 15 GB erhöht. Frühere Versionen haben eine Standardgröße für das Root-Volume von 10 GB.
- Wenn Sie einen Cluster mit der neuesten Patch-Version von Amazon EMR 5.36 oder höher, 6.6 oder höher oder 7.0 oder höher starten, verwendet Amazon EMR die neueste Version von Amazon Linux 2023 oder Amazon Linux 2 für das standardmäßige Amazon EMR-AMI. Weitere Informationen finden Sie unter [Verwenden des standardmäßigen Amazon-Linux-AMI für Amazon EMR](#).

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 223.0	4,14.336	8. März 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost), China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024131.0	4,14.336	14. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024124.0	4,14.336	7. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024109.0	4,14.334	24. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 218.0	4,14.330	2. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202306.0	4,14.330	22. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 116.0	4,14.328	11. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023101.0	4,14.327	16. November 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023020.1	4,14.326	7. November 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsReleaseLabel (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023012.1	4,14.326	26. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 926.0	4,14.322	19. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)
2.0.2023 906.0	4,14.322	04. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 822.0	4,14.322	30. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.20230808.0	4,14.320	24. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv)



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 727.0	4,14.320	14. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 719,0	4,14.320	02. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 628.0	4,14.318	12. Juli 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral)

## Version 6.11.1

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 6.11.1. Änderungen beziehen sich auf Version 6.11.0. Informationen zum Zeitplan für die Veröffentlichung finden Sie unter [6.11.1 Änderungsprotokoll](#).

### Änderungen, Verbesserungen und behobene Probleme

- Aufgrund von Sperrkonflikten kann ein Knoten in einen Deadlock geraten, wenn er gleichzeitig mit dem Versuch, ihn außer Betrieb zu nehmen, hinzugefügt oder entfernt wird. Infolgedessen reagiert der Hadoop Resource Manager (YARN) nicht mehr und wirkt sich auf alle eingehenden und aktuell laufenden Container aus.
- Diese Version enthält eine Änderung, die es Hochverfügbarkeitsclustern ermöglicht, nach einem Neustart den Status „Fehler“ wiederherzustellen.
- Diese Version enthält Sicherheitsupdates für Hue und HBase.
- Diese Version behebt ein Problem, bei dem Cluster, die Workloads auf Spark mit Amazon EMR ausführen, möglicherweise unbemerkt falsche Ergebnisse mit `contains`, `startsWith`, `endsWith` und `like` erhalten. Dieses Problem tritt auf, wenn Sie die Ausdrücke für partitionierte Felder verwenden, die Metadaten im Amazon EMR Hive3 Metastore Server (HMS) haben.
- Diese Version behebt ein Problem mit der Drosselung auf der Glue-Seite, wenn es keine benutzerdefinierten Funktionen (UDF) gibt.
- Diese Version behebt ein Problem, bei dem Container-Protokolle vom Knotenprotokoll-Aggregationservice gelöscht werden, bevor Log Pusher sie im Falle einer Außerbetriebnahme von YARN an S3 weiterleiten kann.
- Diese Version behebt ein Problem mit FairShare Scheduler-Metriken, wenn Node Label für Hadoop aktiviert ist.
- Diese Version behebt ein Problem, das die Leistung von Spark beeinträchtigte, wenn Sie einen `true`-Standardwert für die `spark.yarn.heterogeneousExecutors.enabled`-Konfiguration in `spark-defaults.conf` festlegen.
- Diese Version behebt ein Problem, bei dem Reduce Task die Shuffle-Daten nicht lesen konnte. Das Problem führte zu Hive-Abfragefehlern mit einem Speicherfehler.
- Diese Version fügt dem Cluster-Skalierungs-Workflow für EMR-Cluster, auf denen Presto oder Trino ausgeführt werden, einen neuen Wiederholungsmechanismus hinzu. Diese Verbesserung verringert das Risiko, dass die Clustergrößenänderung aufgrund eines einzigen fehlgeschlagenen Größenänderungsvorgangs auf unbestimmte Zeit zum Stillstand kommt. Sie verbessert auch die Clusterauslastung, da Ihr Cluster schneller hoch- und herunterskaliert wird.

- Diese Version verbessert die Cluster-Herunterskalierungs-Logik, sodass Ihr Cluster nicht versucht, die Core-Knoten unter die für den Cluster eingestellte HDFS-Replikationsfaktor-Einstellung herunterzuskalieren. Dies entspricht Ihren Anforderungen an die Datenredundanz und verringert die Wahrscheinlichkeit, dass ein Skalierungsvorgang zum Stillstand kommt.
- Der Protokoll-Management-Daemon wurde aktualisiert, um alle Protokolle, die aktiv verwendet werden, mit offenen Datei-Handles auf dem lokalen Instance-Speicher und die zugehörigen Prozesse zu identifizieren. Dieses Upgrade stellt sicher, dass Amazon EMR die Dateien ordnungsgemäß löscht und Speicherplatz zurückgewinnt, nachdem die Protokolle in Amazon S3 archiviert wurden.
- Diese Version beinhaltet eine Erweiterung des Protokollverwaltungs-Daemons, die leere, unbenutzte Schrittverzeichnisse im lokalen Cluster-Dateisystem löscht. Eine zu große Anzahl leerer Verzeichnisse kann die Leistung der Amazon-EMR-Daemons beeinträchtigen und zu einer Überauslastung der Festplatte führen.
- Wenn Sie einen Cluster mit der neuesten Patch-Version von Amazon EMR 5.36 oder höher, 6.6 oder höher oder 7.0 oder höher starten, verwendet Amazon EMR die neueste Version von Amazon Linux 2023 oder Amazon Linux 2 für das standardmäßige Amazon EMR-AMI. Weitere Informationen finden Sie unter [Verwenden des standardmäßigen Amazon-Linux-AMI für Amazon EMR](#).

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 223.0	4,14.336	8. März 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost), China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024131.0	4,14.336	14. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024124.0	4,14.336	7. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024109.0	4,14.334	24. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia), Kanada West (Calgary)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 218.0	4,14.330	2. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202306.0	4,14.330	22. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsReleaseLabel (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023116.0	4,14.328	11. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsReleaseLabel (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023101.0	4,14.327	16. November 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023020.1	4,14.326	7. November 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023012.1	4,14.326	26. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 926.0	4,14.322	19. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv), AWS GovCloud (US-West), AWS GovCloud (US-Ost),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			China (Peking), China (Ningxia)
2.0.2023 906.0	4,14.322	04. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 822.0	4,14.322	30. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.20230808.0	4,14.320	24. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023.0727.0	4,14.320	14. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral)

## Version 6.11.0

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 6.11.0. Änderungen beziehen sich auf Version 6.10.0. Informationen zum Zeitplan der Veröffentlichung finden Sie im [Änderungsprotokoll](#).

## Neue Features

- Amazon EMR 6.11.0 unterstützt Apache Spark 3.3.2-amzn-0, Apache Spark RAPIDS 23.02.0-amzn-0, CUDA 11.8.0, Apache Hudi 0.13.0-amzn-0, Apache Iceberg 1.2.0-amzn-0, Trino 410-amzn-0 und PrestoDB 0.279-amzn-0.

## Änderungen, Verbesserungen und behobene Probleme

- Mit Amazon EMR 6.11.0 wurde der DynamoDB-Konnektor auf Version 5.0.0 aktualisiert. Version 5.0.0 verwendet. AWS SDK for Java 2.x In früheren Versionen wurde AWS SDK for Java 1.x verwendet. Aufgrund dieses Upgrades empfehlen wir Ihnen dringend, Ihren Code zu testen, bevor Sie den DynamoDB-Konnektor mit Amazon EMR 6.11 verwenden.
- Wenn der DynamoDB-Konnektor für Amazon EMR 6.11.0 den DynamoDB-Service aufruft, verwendet er den Region-Wert, den Sie für die Eigenschaft `dynamodb.endpoint` angeben. Wir empfehlen, dass Sie auch `dynamodb.region` konfigurieren, wenn Sie `dynamodb.endpoint` verwenden, und dass beide Eigenschaften auf dieselbe AWS-Region abzielen. Wenn Sie verwenden `dynamodb.endpoint` und nicht konfigurierend `dynamodb.region`, gibt der DynamoDB-Connector für Amazon EMR 6.11.0 eine ungültige Region-Ausnahme zurück und versucht, Ihre AWS-Region Informationen aus dem Amazon EC2 EC2-Instance-Metadaten-Service (IMDS) abzugleichen. Wenn der Konnektor die Region nicht aus IMDS abrufen kann, verwendet er standardmäßig USA Ost (Nord-Virginia) (`us-east-1`). Der folgende Fehler ist ein Beispiel für die ungültige Region-Ausnahme, die möglicherweise angezeigt wird, wenn Sie die `dynamodb.region` Eigenschaft nicht richtig konfigurieren: `error software.amazon.awssdk.services.dynamodb.model.DynamoDbException: Credential should be scoped to a valid region`. Weitere Informationen zu den Klassen, die vom AWS SDK for Java Upgrade auf 2.x betroffen sind, finden Sie im Commit [Upgrade AWS SDK for Java from 1.x auf 2.x \(#175\)](#) im GitHub Repository für den Amazon EMR — DynamoDB-Connector.
- Diese Version behebt ein Problem, bei dem Spaltendaten zu NULL werden, wenn Sie Delta Lake zum Speichern von Delta-Tabellendaten in Amazon S3 nach dem Spaltenumbenennungsvorgang verwenden. Weitere Informationen zu diesem experimentellen Feature in Delta Lake finden Sie unter [Vorgang zum Umbenennen von Spalten](#) im Delta-Lake-Benutzerhandbuch.
- Die Version 6.11.0 behebt ein Problem, das auftreten kann, wenn Sie einen Edge-Knoten erstellen, indem Sie einen der Primärknoten aus einem Cluster mit mehreren Primärknoten replizieren. Der replizierte Edge-Knoten kann zu Verzögerungen bei Scale-Down-Vorgängen oder zu einer hohen Arbeitsspeicherauslastung auf den Primärknoten führen. Weitere Informationen zum Erstellen


eines Edge-Knotens für die Kommunikation mit Ihrem EMR-Cluster finden Sie unter [Edge Node Creator](#) im `aws-samples` Repo unter GitHub

- Die Version 6.11.0 verbessert den Automatisierungsprozess, den Amazon EMR verwendet, um Amazon-EBS-Volumes nach einem Neustart erneut in eine Instance einzubinden.
- Die Version 6.11.0 behebt ein Problem, das zu zeitweiligen Lücken in den Hadoop-Metriken führte, die Amazon EMR auf Amazon veröffentlicht. CloudWatch
- Die Version 6.11.0 behebt ein Problem mit EMR-Clustern, bei dem ein Update der YARN-Konfigurationsdatei, die die Ausschlussliste der Knoten für den Cluster enthält, aufgrund einer Überauslastung der Festplatte unterbrochen wird. Das unvollständige Update behindert zukünftige Cluster-Herunterskalierungs-Vorgänge. Diese Version stellt sicher, dass Ihr Cluster fehlerfrei bleibt und dass die Skalierungsvorgänge wie erwartet funktionieren.
- Die Standardgröße des Root-Volumes wurde in Amazon EMR 6.10.0 und höher auf 15 GB erhöht. Frühere Versionen haben eine Standardgröße für das Root-Volume von 10 GB.
- Mit Hadoop 3.3.3 wurde eine Änderung in YARN ([YARN-9608](#)) eingeführt, die dafür sorgt, dass Knoten, auf denen Container laufen, außer Betrieb genommen werden, bis die Anwendung abgeschlossen ist. Diese Änderung stellt sicher, dass lokale Daten wie Shuffle-Daten nicht verloren gehen und Sie den Auftrag nicht erneut ausführen müssen. Dieser Ansatz kann auch zu einer Unterauslastung von Ressourcen in Clustern mit oder ohne aktivierter verwalteter Skalierung führen.

Bei den Amazon-EMR-Versionen 6.11.0 und höher sowie 6.8.1, 6.9.1 und 6.10.1 ist der Wert in `yarn-site.xml` von `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` auf `false` gesetzt, um dieses Problem zu beheben.

Der Fix behebt zwar die Probleme, die durch YARN-9608 eingeführt wurden, kann jedoch dazu führen, dass Hive-Aufträge aufgrund von Shuffle-Datenverlusten auf Clustern, für die verwaltete Skalierung aktiviert ist, fehlschlagen. Wir haben dieses Risiko in dieser Version verringert, indem wir auch `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-shuffle-data` für Hive-Workloads eingerichtet haben. Diese Konfiguration ist nur mit Amazon-EMR-Versionen 6.11.0 und höher verfügbar.

- Wenn Sie einen Cluster mit der neuesten Patch-Version von Amazon EMR 5.36 oder höher, 6.6 oder höher oder 7.0 oder höher starten, verwendet Amazon EMR die neueste Version von Amazon Linux 2023 oder Amazon Linux 2 für das standardmäßige Amazon EMR-AMI. Weitere Informationen finden Sie unter [Verwenden des standardmäßigen Amazon-Linux-AMI für Amazon EMR](#).

 Note

Diese Version erhält keine automatischen AMI-Updates mehr, da sie von einer weiteren Patch-Versionen abgelöst wurde. Die Patch-Version wird durch die Zahl nach dem zweiten Dezimalpunkt (6.8.**1**) gekennzeichnet. Um zu sehen, ob Sie die neueste Patch-Version verwenden, überprüfen Sie die verfügbaren Versionen im [Versionshandbuch](#) oder überprüfen Sie die Amazon-EMR-Versions-Dropdownliste, wenn Sie einen Cluster in der Konsole erstellen, oder verwenden Sie die [ListReleaseLabels](#)-API- oder [list-release-labels](#)-CLI-Aktion. Um aktuelle Informationen über Neuerscheinungen erhalten, abonnieren Sie den RSS-Feed auf der Seite [Was ist neu?](#).



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.20230808.0	4,14.320	24. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 727.0	4,14.320	14. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 719,0	4,14.320	02. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 628.0	4,14.318	12. Juli 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Zürich), Europa (Mailand), Europa (Spanien), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 612,0	4,14.314	23. Juni 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Zürich), Europa (Mailand), Europa (Spanien), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 504.1	4,14.313	16. Mai 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral)

## Version 6.10.0

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 6.10.0. Änderungen beziehen sich auf Version 6.9.0. Informationen zum Zeitplan der Veröffentlichung finden Sie im [Änderungsprotokoll](#).

### Neue Features

- Amazon EMR 6.10.0 unterstützt Apache Spark 3.3.1, Apache Spark RAPIDS 22.12.0, CUDA 11.8.0, Apache Hudi 0.12.2-amzn-0, Apache Iceberg 1.1.0-amzn-0, Trino 403 und PrestoDB 0.278.1.
- Amazon EMR 6.10.0 enthält einen nativen Trino-Hudi-Konnektor, der Lesezugriff auf Daten in Hudi-Tabellen ermöglicht. Sie können den Konnektor mit `trino-cli --catalog hudi` aktivieren und den Konnektor für Ihre Anforderungen mit `trino-connector-hudi` konfigurieren. Durch die native Integration mit Amazon EMR müssen Sie `trino-connector-hive` nicht mehr zum Abfragen von Hudi-Tabellen verwenden. Eine Liste der unterstützten Konfigurationen mit dem neuen Konnektor finden Sie auf der [Hudi-Konnektor-Seite](#) der Trino-Dokumentation.
- Amazon-EMR-Versionen 6.10.0 und höher unterstützen die Apache-Zeppelin-Integration mit Apache Flink. Weitere Informationen finden Sie unter [Arbeiten mit Flink-Aufträgen von Zeppelin in Amazon EMR](#).

### Bekannte Probleme


- Mit Hadoop 3.3.3 wurde eine Änderung in YARN ([YARN-9608](#)) eingeführt, die dafür sorgt, dass Knoten, auf denen Container laufen, außer Betrieb genommen werden, bis die Anwendung abgeschlossen ist. Diese Änderung stellt sicher, dass lokale Daten wie Shuffle-Daten nicht verloren gehen und Sie den Auftrag nicht erneut ausführen müssen. Dieser Ansatz kann auch zu einer Unterauslastung von Ressourcen in Clustern mit oder ohne aktivierter verwalteter Skalierung führen.

Um dieses Problem in Amazon EMR 6.10.0 zu umgehen, können Sie den Wert `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` auf `false` in `yarn-site.xml` setzen. In den Amazon-EMR-Versionen 6.11.0 und höher sowie 6.8.1, 6.9.1 und 6.10.1 ist die Konfiguration standardmäßig auf `false` eingestellt, um dieses Problem zu beheben.

## Änderungen, Verbesserungen und behobene Probleme

- Amazon EMR 6.10.0 entfernt die Abhängigkeit von der `minimal-json.jar` für die [Amazon-Redshift-Integration für Apache Spark](#) und fügt automatisch die erforderlichen Spark-Redshift-bezogenen JAR-Dateien zum Executor-Klassenpfad für Spark hinzu: `spark-redshift.jar`, `spark-avro.jar` und `RedshiftJDBC.jar`.
- Die Version 6.10.0 verbessert den On-Cluster-Log-Management-Daemon zur Überwachung zusätzlicher Protokollordner in Ihrem EMR-Cluster. Durch diese Verbesserung werden Szenarien mit übermäßiger Festplattenauslastung minimiert.
- Die Version 6.10.0 startet den Protokoll-Management-Daemon auf dem Cluster automatisch neu, wenn er beendet wird. Durch diese Verbesserung wird das Risiko verringert, dass Knoten aufgrund übermäßiger Festplattenauslastung als fehlerhaft erscheinen.
- Amazon EMR 6.10.0 unterstützt regionale Endpunkte für die EMRFS-Benutzerzuordnung.
- Die Standardgröße des Root-Volumes wurde in Amazon EMR 6.10.0 und höher auf 15 GB erhöht. Frühere Versionen haben eine Standardgröße für das Root-Volume von 10 GB.
- Die Version 6.10.0 behebt ein Problem, das dazu führte, dass Spark-Aufträge zum Stillstand kamen, wenn sich alle verbleibenden Spark-Ausführern auf einem außer Betrieb genommenen Host mit dem YARN-Ressourcenmanager befanden.
- Bei Amazon EMR 6.6.0 bis 6.9.x haben INSERT-Abfragen mit dynamischer Partition und einer ORDER-BY- oder SORT-BY-Klausel immer zwei Reduzierungen. Dieses Problem wird durch die OSS-Änderung [HIVE-20703](#) verursacht, die die Optimierung dynamischer Sortierpartitionen einer kostenbasierten Entscheidung unterstellt. Wenn Ihr Workload keine Sortierung dynamischer Partitionen erfordert, empfehlen wir Ihnen, die `hive.optimize.sort.dynamic.partition.threshold`-Eigenschaft auf `-1` zu setzen, um das neue Feature zu deaktivieren und die korrekt berechnete Anzahl von Reduzierern zu erhalten. Dieses Problem wurde behoben in OSS Hive als Teil von [HIVE-22269](#) und in Amazon EMR 6.10.0 behoben.
- Wenn Sie einen Cluster mit der neuesten Patch-Version von Amazon EMR 5.36 oder höher, 6.6 oder höher oder 7.0 oder höher starten, verwendet Amazon EMR die neueste Version von Amazon Linux 2023 oder Amazon Linux 2 für das standardmäßige Amazon EMR-AMI. Weitere Informationen finden Sie unter [Verwenden des standardmäßigen Amazon-Linux-AMI für Amazon EMR](#).



 Note

Diese Version erhält keine automatischen AMI-Updates mehr, da sie von einer weiteren Patch-Versionen abgelöst wurde. Die Patch-Version wird durch die Zahl nach dem zweiten Dezimalpunkt (6.8.**1**) gekennzeichnet. Um zu sehen, ob Sie die neueste Patch-Version verwenden, überprüfen Sie die verfügbaren Versionen im [Versionshandbuch](#) oder überprüfen Sie die Amazon-EMR-Versions-Dropdownliste, wenn Sie einen Cluster in der Konsole erstellen, oder verwenden Sie die [ListReleaseLabels](#)-API- oder [list-release-labels](#)-CLI-Aktion. Um aktuelle Informationen über Neuerscheinungen erhalten, abonnieren Sie den RSS-Feed auf der Seite [Was ist neu?](#).

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.20230808.0	4,14.320	24. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 727.0	4,14.320	14. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 719,0	4,14.320	02. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 628.0	4,14.318	12. Juli 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Zürich), Europa (Mailand), Europa (Spanien), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 612,0	4,14.314	23. Juni 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Zürich), Europa (Mailand), Europa (Spanien), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 504.1	4,14.313	16. Mai 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Zürich), Europa (Mailand), Europa (Spanien), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 418,0	4,14.311	3. Mai 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Zürich), Europa (Mailand), Europa (Spanien), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE)



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 404.1	4,14.311	18. April 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE)
2.0.2023 404.0	4,14.311	10. April 2023	USA Ost (Nord-Virginia), Europa (Paris)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 320.0	4,14.309	30. März 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 207,0	4,14.304	22. Februar 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE)

## Version 6.9.0

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 6.9.0. Änderungen beziehen sich auf Amazon-EMR-Version 6.8.0. Informationen zum Zeitplan der Veröffentlichung finden Sie im [Änderungsprotokoll](#).

## Neue Features

- Amazon EMR Version 6.9.0 unterstützt Apache Spark RAPIDS 22.08.0, Apache Hudi 0.12.1, Apache Iceberg 0.14.1, Trino 398 und Tez 0.10.2.
- Amazon EMR Version 6.9.0 enthält eine neue Open-Source-Anwendung, [Delta Lake](#) 2.1.0.
- Die Amazon-Redshift-Integration für Apache Spark ist in den Amazon-EMR-Versionen 6.9.0 und höher enthalten. Die native Integration war bisher ein Open-Source-Tool und ist ein Spark-Konnektor, mit dem Sie Apache-Spark-Anwendungen erstellen können, die Daten in Amazon Redshift und Amazon Redshift Serverless lesen und in diese schreiben. Weitere Informationen finden Sie unter [Amazon-Redshift-Integration für Apache Spark mit Amazon EMR verwenden](#).
- Amazon EMR Version 6.9.0 bietet Unterstützung für die Archivierung von Protokollen in Amazon S3 während der Cluster-Herunterskalierung. Bisher konnten Sie Protokolldateien nur während der Clusterbeendigung in Amazon S3 archivieren. Die neue Funktion stellt sicher, dass auf dem Cluster generierte Protokolldateien auch nach dem Beenden des Knotens auf Amazon S3 bestehen bleiben. Weitere Informationen finden Sie unter [Konfigurieren der Cluster-Protokollierung und des Debuggings](#).
- Um Abfragen mit langer Laufzeit zu unterstützen, verfügt Trino jetzt über einen fehlertoleranten Ausführungsmechanismus. Die fehlertolerante Ausführung minimiert Abfragefehler, indem fehlgeschlagene Abfragen oder deren Komponentenaufgaben wiederholt werden. Weitere Informationen finden Sie unter [Fehlertolerante Ausführung in Trino](#).
- Sie können Apache Flink auf Amazon EMR für die einheitliche BATCH- und STREAM-Verarbeitung von Apache-Hive-Tabellen oder Metadaten beliebiger Flink-Tabellenquellen wie Iceberg, Kinesis oder Kafka verwenden. Sie können den AWS Glue-Datenkatalog mithilfe der Amazon EMR-API, der AWS Management Console, der AWS CLI, oder als Metastore für Flink angeben. Weitere Informationen finden Sie unter [Konfiguration von Flink in Amazon EMR](#).
- Sie können jetzt AWS Identity and Access Management (IAM-) Runtime-Rollen und eine AWS Lake Formation basierte Zugriffskontrolle für Apache Spark-, Apache Hive- und Presto-Abfragen auf Amazon EMR auf EC2-Clustern mit Amazon Studio angeben. SageMaker Weitere Informationen finden Sie unter [Laufzeit-Rollen für Amazon EMR konfigurieren](#).

## Bekannte Probleme

- Für Amazon-EMR-Version 6.9.0 funktioniert Trino nicht auf Clustern, die für Apache Ranger aktiviert sind. Wenn Sie Trino mit Ranger verwenden müssen, wenden Sie sich an [AWS Support](#).

- Wenn Sie die Amazon-Redshift-Integration für Apache Spark verwenden und eine Zeit, `timetz`, `timestamp` oder `timestampz` mit Mikrosekundengenauigkeit im Parquet-Format haben, rundet der Konnektor die Zeitwerte auf den nächstliegenden Millisekundenwert. Um das Problem zu umgehen, verwenden Sie den `unload_s3_format`-Formatparameter-Text-Unload.
- Wenn Sie Spark mit der Formatierung des Hive-Partitionsstandorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den Amazon-EMR-Versionen 5.30.0 bis 5.36.0 und 6.2.0 bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:
  - Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.
  - Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist `s3://bucket/table/p=a` ein Präfix von `s3://bucket/table/p=a b`.
  - Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen UTF-8-Wert, der kleiner als das `/`-Zeichen (U+002F) ist. Beispielsweise fällt das Leerzeichen (U+0020), das in `s3://bucket/table/p=a b` zwischen `a` und `b` vorkommt, in diese Kategorie. Beachten Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: `!"#$%&'()*+,-.` Weitere Informationen finden Sie unter [UTF-8-Kodierungstabelle und Unicode-Zeichen](#).

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.

- Verbindungen zu Amazon EMR-Clustern von Amazon SageMaker Studio aus können zeitweise fehlschlagen und der Antwortcode 403 Forbidden wird angezeigt. Dieser Fehler tritt auf, wenn die Einrichtung der IAM-Rolle auf dem Cluster länger als 60 Sekunden dauert. Um dieses Problem zu umgehen, können Sie einen Amazon-EMR-Patch installieren, um Wiederholungsversuche zu ermöglichen und das Timeout auf mindestens 300 Sekunden zu erhöhen. Verwenden Sie die folgenden Schritte, um die Bootstrap-Aktion anzuwenden, wenn Sie Ihren Cluster starten.

1. Laden Sie das Bootstrap-Skript und die RPM-Dateien von den folgenden Amazon-S3-URIs herunter.

```
s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/gcsc/replace-rpms.sh
s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/gcsc/emr-secret-agent-1.18.0-SNAPSHOT20221121212949.noarch.rpm
```

2. Laden Sie die Dateien aus dem vorherigen Schritt in einen Amazon-S3-Bucket hoch, dessen Eigentümer Sie sind. Der Bucket muss sich in dem Bereich befinden, in AWS-Region dem Sie den Cluster starten möchten.
3. Schließen Sie die folgende Bootstrap-Aktion ein, wenn Sie Ihren EMR-Cluster starten. Ersetzen Sie *bootstrap\_URI* und *RPM\_URI* durch die entsprechenden URIs von Amazon S3.

```
--bootstrap-actions "Path=bootstrap_URI,Args=[RPM_URI]"
```

- Bei den Amazon-EMR-Versionen 5.36.0 und 6.6.0 bis 6.9.0 kann es bei SecretAgent- und RecordServer-Servicekomponenten aufgrund einer falschen Konfiguration des Dateinamenmusters in den Log4j2-Eigenschaften zu einem Verlust von Protokolldaten kommen. Die falsche Konfiguration führt dazu, dass die Komponenten nur eine Protokolldatei pro Tag generieren. Wenn die Rotationsstrategie angewendet wird, überschreibt sie die vorhandene Datei, anstatt wie erwartet eine neue Protokolldatei zu generieren. Um das Problem zu umgehen, verwenden Sie eine Bootstrap-Aktion, um jede Stunde Protokolldateien zu generieren, und fügen Sie eine automatisch inkrementierte Ganzzahl an den Dateinamen an, um die Rotation zu handhaben.

Verwenden Sie für die Versionen Amazon EMR 6.6.0 bis 6.9.0 die folgende Bootstrap-Aktion, wenn Sie einen Cluster starten.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-6x/replace-puppet.sh,Args=[]"
```

Verwenden Sie für Amazon EMR 5.36.0 die folgende Bootstrap-Aktion, wenn Sie einen Cluster starten.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-5x/replace-puppet.sh,Args=[]"
```

- Apache Flink bietet native S3 FileSystem - und FileSystem Hadoop-Konnektoren, mit denen Anwendungen Daten erstellen FileSink und in Amazon S3 schreiben können. Dies FileSink schlägt mit einer der folgenden zwei Ausnahmen fehl.

```
java.lang.UnsupportedOperationException: Recoverable writers on Hadoop are only supported for HDFS
```

```
Caused by: java.lang.NoSuchMethodError:
  org.apache.hadoop.io.retry.RetryPolicies.retryOtherThanRemoteAndSaslException(Lorg/
  apache/hadoop/io/retry/RetryPolicy;Ljava/util/Map;)Lorg/apache/hadoop/io/retry/
  RetryPolicy;
                                     at
  org.apache.hadoop.yarn.client.RMProxy.createRetryPolicy(RMProxy.java:302) ~[hadoop-
  yarn-common-3.3.3-amzn-0.jar:?]
```

Als Workaround können Sie einen Amazon-EMR-Patch installieren, der das oben genannte Problem in Flink behebt. Führen Sie die folgenden Schritte aus, um die Bootstrap-Aktion beim Start Ihres Clusters anzuwenden.

1. Laden Sie das flink-rpm in Ihren Amazon-S3-Bucket herunter. Ihr RPM-Pfad ist `s3://DOC-EXAMPLE-BUCKET/rpms/flink/`.
2. Laden Sie das Bootstrap-Skript und die RPM-Dateien mit dem folgenden URI von Amazon S3 herunter. `regionName` Ersetzen Sie durch den AWS-Region Ort, an dem Sie den Cluster starten möchten.

```
s3://emr-data-access-control-regionName/customer-bootstrap-actions/gcsc/replace-
rpms.sh
```

3. Mit Hadoop 3.3.3 wurde eine Änderung in YARN ([YARN-9608](#)) eingeführt, die dafür sorgt, dass Knoten, auf denen Container laufen, außer Betrieb genommen werden, bis die Anwendung abgeschlossen ist. Diese Änderung stellt sicher, dass lokale Daten wie Shuffle-Daten nicht verloren gehen und Sie den Auftrag nicht erneut ausführen müssen. In Amazon EMR 6.8.0 und 6.9.0 kann dieser Ansatz auch zu einer Unterauslastung von Ressourcen auf Clustern mit oder ohne aktivierter verwalteter Skalierung führen.

Mit [Amazon EMR 6.10.0](#) gibt es einen Workaround für dieses Problem, indem Sie den Wert in `yarn-site.xml` von `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` auf `false` setzen. In den Amazon-EMR-Versionen 6.11.0 und höher sowie 6.8.1, 6.9.1 und 6.10.1 ist die Konfiguration standardmäßig auf `false` eingestellt, um dieses Problem zu beheben.

## Änderungen, Verbesserungen und behobene Probleme

- Für Amazon-EMR-Version 6.9.0 und höher verwenden alle von Amazon EMR installierten Komponenten, die Log4j-Bibliotheken verwenden, Log4j Version 2.17.1 oder höher.

- Wenn Sie den DynamoDB-Konnektor mit Spark auf den Amazon-EMR-Versionen 6.6.0, 6.7.0 und 6.8.0 verwenden, geben alle Lesevorgänge aus Ihrer Tabelle ein leeres Ergebnis zurück, obwohl der Eingabe-Split auf nicht leere Daten verweist. Amazon EMR Version 6.9.0 behebt dieses Problem.
- Amazon EMR 6.9.0 bietet eingeschränkte Unterstützung für die auf Lake Formation basierende Zugriffskontrolle mit Apache Hudi beim Lesen von Daten mit Spark SQL. Die Unterstützung gilt für SELECT-Abfragen mit Spark SQL und ist auf die Zugriffskontrolle auf Spaltenebene beschränkt. Weitere Informationen finden Sie unter [Hudi und Lake Formation](#).
- Wenn Sie Amazon EMR 6.9.0 verwenden, um einen Hadoop-Cluster mit aktivierten [Knoten-Kennzeichnungen](#) zu erstellen, gibt die [YARN-Metriken-API](#) aggregierte Informationen über alle Partitionen statt der Standardpartition zurück. Weitere Informationen finden Sie unter [YARN-11414](#).
- Mit Amazon-EMR-Version 6.9.0 haben wir Trino auf Version 398 aktualisiert, die Java 17 verwendet. Die vorherige unterstützte Version von Trino für Amazon EMR 6.8.0 war Trino 388, das auf Java 11 lief. Weitere Informationen zu dieser Änderung finden Sie unter [Trino-Updates für Java 17 im Trino-Blog](#).
- Diese Version behebt ein Problem mit der Nichtübereinstimmung der Timing-Sequenz zwischen Apache BigTop und Amazon EMR in der EC2-Cluster-Startsequenz. Diese Nichtübereinstimmung der Timing-Sequenz tritt auf, wenn ein System versucht, zwei oder mehr Operationen gleichzeitig auszuführen, anstatt sie in der richtigen Reihenfolge auszuführen. Infolgedessen kam es bei bestimmten Clusterkonfigurationen zu Timeouts beim Instance-Start und zu langsameren Cluster-Startup-Zeiten.
- Wenn Sie einen Cluster mit der neuesten Patch-Version von Amazon EMR 5.36 oder höher, 6.6 oder höher oder 7.0 oder höher starten, verwendet Amazon EMR die neueste Version von Amazon Linux 2023 oder Amazon Linux 2 für das standardmäßige Amazon EMR-AMI. Weitere Informationen finden Sie unter [Verwenden des standardmäßigen Amazon-Linux-AMI für Amazon EMR](#).

#### Note

Diese Version erhält keine automatischen AMI-Updates mehr, da sie von einer weiteren Patch-Versionen abgelöst wurde. Die Patch-Version wird durch die Zahl nach dem zweiten Dezimalpunkt (6.8.**1**) gekennzeichnet. Um zu sehen, ob Sie die neueste Patch-Version verwenden, überprüfen Sie die verfügbaren Versionen im [Versionshandbuch](#) oder überprüfen Sie die Amazon-EMR-Versions-Dropdownliste, wenn Sie einen Cluster in der Konsole erstellen, oder verwenden Sie die [ListReleaseLabels](#)-API- oder [list-](#)



[release-labels](#)-CLI-Aktion. Um aktuelle Informationen über Neuerscheinungen erhalten, abonnieren Sie den RSS-Feed auf der Seite [Was ist neu?](#).

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 808.0	4,14.320	24. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 727.0	4,14.320	14. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 719,0	4,14.320	02. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral), Israel (Tel Aviv)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 628.0	4,14.318	12. Juli 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 612,0	4,14.314	23. Juni 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.20230414.1	4.14.313	16. Mai 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 418,0	4,14.311	3. Mai 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 404.1	4,14.311	18. April 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)
2.0.2023 404.0	4,14.311	10. April 2023	USA Ost (Nord-Virginia), Europa (Paris)



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 320.0	4,14.309	30. März 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 307,0	4,14.305	15. März 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202307,0	4,14.304	22. Februar 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022 210.1	4.14.301	12. Januar 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022103.3	4,14.296	5. Dezember 2022	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

## Version 6.8.0

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 6.8.0. Änderungen beziehen sich auf Version 6.7.0.

## Neue Features

- Das Amazon-EMR-Schrittfeature unterstützt jetzt Apache.Livy-Endpunkt- und JDBC/ODBC-Clients. Weitere Informationen finden Sie unter [Laufzeit-Rollen für Amazon EMR konfigurieren](#).
- Amazon-EMR-Version 6.8.0 wird mit Apache-HBase-Version 2.4.12 geliefert. Mit dieser HBase-Version können Sie Ihre HBase-Tabellen sowohl archivieren als auch löschen. Der Amazon-S3-Archivierungsprozess benennt alle Tabellendateien in das Archivverzeichnis um. Dies kann ein kostspieliger und langwieriger Prozess sein. Jetzt können Sie den Archivierungsvorgang überspringen und große Tabellen schnell löschen und löschen. Weitere Informationen finden Sie unter [Verwenden der HBase-Shell](#).

## Bekannte Probleme

- Mit Hadoop 3.3.3 wurde eine Änderung in YARN ([YARN-9608](#)) eingeführt, die dafür sorgt, dass Knoten, auf denen Container laufen, außer Betrieb genommen werden, bis die Anwendung abgeschlossen ist. Diese Änderung stellt sicher, dass lokale Daten wie Shuffle-Daten nicht verloren gehen und Sie den Auftrag nicht erneut ausführen müssen. In Amazon EMR 6.8.0 und 6.9.0 kann dieser Ansatz auch zu einer Unterauslastung von Ressourcen auf Clustern mit oder ohne aktivierter verwalteter Skalierung führen.

Mit [Amazon EMR 6.10.0](#) gibt es einen Workaround für dieses Problem, indem Sie den Wert in `yarn-site.xml` von `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` auf `false` setzen. In den Amazon-EMR-Versionen 6.11.0 und höher sowie 6.8.1, 6.9.1 und 6.10.1 ist die Konfiguration standardmäßig auf `false` eingestellt, um dieses Problem zu beheben.

## Änderungen, Verbesserungen und behobene Probleme

- Als Amazon-EMR-Version 6.5.0, 6.6.0 oder 6.7.0 Apache Phoenix-Tabellen über die Apache Spark-Shell las, erzeugte Amazon EMR eine `NoSuchMethodError`. Amazon EMR Version 6.8.0 behebt dieses Problem.
- Amazon-EMR-Version 6.8.0 enthält [Apache Hudi](#) 0.11.1. Amazon-EMR-6.8.0-Cluster sind jedoch auch mit der Open-Source-Version von Hudi 0.12.0 kompatibel. `hudi-spark3.3-bundle_2.12`
- Amazon-EMR-Version 6.8.0 wird mit Apache Spark 3.3.0 geliefert. Diese Spark-Version verwendet Apache Log4j 2 und die `log4j2.properties`-Datei zur Konfiguration von Log4j in Spark-Prozessen. Wenn Sie Spark im Cluster verwenden oder EMR-Cluster mit benutzerdefinierten

Konfigurationsparametern erstellen und ein Upgrade auf Amazon-EMR-Version 6.8.0 durchführen möchten, müssen Sie auf die neue `spark-log4j2`-Konfigurationsklassifizierung und das neue Schlüsselformat für Apache Log4j 2 migrieren. Weitere Informationen finden Sie unter [Migration von Apache Log4j 1.x zu Log4j 2.x](#).

- Wenn Sie einen Cluster mit der neuesten Patch-Version von Amazon EMR 5.36 oder höher, 6.6 oder höher oder 7.0 oder höher starten, verwendet Amazon EMR die neueste Version von Amazon Linux 2023 oder Amazon Linux 2 für das standardmäßige Amazon EMR-AMI. Weitere Informationen finden Sie unter [Verwenden des standardmäßigen Amazon-Linux-AMI für Amazon EMR](#).

### Note

Diese Version erhält keine automatischen AMI-Updates mehr, da sie von einer weiteren Patch-Versionen abgelöst wurde. Die Patch-Version wird durch die Zahl nach dem zweiten Dezimalpunkt (6.8.**1**) gekennzeichnet. Um zu sehen, ob Sie die neueste Patch-Version verwenden, überprüfen Sie die verfügbaren Versionen im [Versionshandbuch](#) oder überprüfen Sie die Amazon-EMR-Versions-Dropdownliste, wenn Sie einen Cluster in der Konsole erstellen, oder verwenden Sie die [ListReleaseLabels](#)-API- oder [list-release-labels](#)-CLI-Aktion. Um aktuelle Informationen über Neuerscheinungen erhalten, abonnieren Sie den RSS-Feed auf der Seite [Was ist neu?](#).

OsReleaseLabel (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 808.0	4,14.320	24. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Asien-Pazifik (Hongkong , Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			Asien-Pazifik (Seoul), Asien Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral)



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 727.0	4,14.320	14. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 719,0	4,14.320	02. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Asien-Pazifik (Melbourne), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 628.0	4,14.318	12. Juli 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 612,0	4,14.314	23. Juni 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.20230414.1	4.14.313	16. Mai 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 418,0	4,14.311	3. Mai 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 404.1	4,14.311	18. April 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)
2.0.2023 404.0	4,14.311	10. April 2023	USA Ost (Nord-Virginia), Europa (Paris)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.20230320.0	4.14.309	30. März 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 307,0	4,14.305	15. März 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202307,0	4,14.304	22. Februar 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023119.1	4,14.301	3. Februar 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022 210.1	4.14.301	22. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022 103.3	4,14.296	5. Dezember 2022	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022004,0	4,14.294	02. November 2022	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022 912.1	4,14.291	6. September 2022	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

## Bekannte Probleme

- Wenn Sie den DynamoDB-Konnektor mit Spark auf den Amazon-EMR-Versionen 6.6.0, 6.7.0 und 6.8.0 verwenden, geben alle Lesevorgänge aus Ihrer Tabelle ein leeres Ergebnis zurück, obwohl der Eingabe-Split auf nicht leere Daten verweist. Das liegt daran, dass Spark 3.2.0 standardmäßig `spark.hadoopRDD.ignoreEmptySplits` auf `true` einstellt. Um das Problem zu umgehen,

setzen Sie explizit `spark.hadoopRDD.ignoreEmptySplits` auf `false`. Amazon EMR Version 6.9.0 behebt dieses Problem.

- Wenn Sie Spark mit der Formatierung des Hive-Partitionsstandorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den Amazon-EMR-Versionen 5.30.0 bis 5.36.0 und 6.2.0 bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:
  - Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.
  - Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist `s3://bucket/table/p=a` ein Präfix von `s3://bucket/table/p=a b`.
  - Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen UTF-8-Wert, der kleiner als das `/`-Zeichen (U+002F) ist. Beispielsweise fällt das Leerzeichen (U+0020), das in `s3://bucket/table/p=a b` zwischen `a` und `b` vorkommt, in diese Kategorie. Beachten Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: `!"#$%&'()*+,-.` Weitere Informationen finden Sie unter [UTF-8-Kodierungstabelle und Unicode-Zeichen](#).

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.

- Bei den Amazon-EMR-Versionen 5.36.0 und 6.6.0 bis 6.9.0 kann es bei `SecretAgent`- und `RecordServer`-Servicekomponenten aufgrund einer falschen Konfiguration des Dateinamenmusters in den `Log4j2`-Eigenschaften zu einem Verlust von Protokolldaten kommen. Die falsche Konfiguration führt dazu, dass die Komponenten nur eine Protokolldatei pro Tag generieren. Wenn die Rotationsstrategie angewendet wird, überschreibt sie die vorhandene Datei, anstatt wie erwartet eine neue Protokolldatei zu generieren. Um das Problem zu umgehen, verwenden Sie eine Bootstrap-Aktion, um jede Stunde Protokolldateien zu generieren, und fügen Sie eine automatisch inkrementierte Ganzzahl an den Dateinamen an, um die Rotation zu handhaben.

Verwenden Sie für die Versionen Amazon EMR 6.6.0 bis 6.9.0 die folgende Bootstrap-Aktion, wenn Sie einen Cluster starten.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-6x/replace-puppet.sh,Args=[]"
```



Verwenden Sie für Amazon EMR 5.36.0 die folgende Bootstrap-Aktion, wenn Sie einen Cluster starten.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-5x/replace-puppet.sh,Args=[]"
```

Informationen zum Zeitplan der Veröffentlichung finden Sie im [Änderungsprotokoll](#).

## Version 6.7.0

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 6.7.0.

Änderungen beziehen sich auf Version 6.6.0.

Erste Version: 15. Juli 2022

### Neue Features

- Amazon EMR unterstützt jetzt Apache Spark 3.2.1, Apache Hive 3.1.3, HUDI 0.11, PrestoDB 0.272 und Trino 0.378.
- Unterstützt Zugriffskontrollen auf Basis der IAM-Rolle und Lake Formation mit EMR-Schritten (Spark, Hive) für Amazon EMR auf EC2-Clustern.
- Unterstützt Apache-Spark-Datendefinitionsanweisungen auf Apache-Ranger-fähigen Clustern. Dies beinhaltet jetzt die Unterstützung für Trino-Anwendungen, die Apache-Hive-Metadaten auf Apache-Ranger-fähigen Clustern lesen und schreiben. Weitere Informationen finden Sie unter [Aktivieren von Verbund-Governance mit Trino und Apache Ranger auf Amazon EMR](#).
- Wenn Sie einen Cluster mit der neuesten Patch-Version von Amazon EMR 5.36 oder höher, 6.6 oder höher oder 7.0 oder höher starten, verwendet Amazon EMR die neueste Version von Amazon Linux 2023 oder Amazon Linux 2 für das standardmäßige Amazon EMR-AMI. Weitere Informationen finden Sie unter [Verwenden des standardmäßigen Amazon-Linux-AMI für Amazon EMR](#).

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 223.0	4,14.336	8. März 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)
2.0.2024 131,0	4,14.336	14. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral), AWS GovCloud (US-West), AWS GovCloud (US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 124,0	4,14.336	7. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral), AWS GovCloud (US-West), AWS GovCloud (US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024109,0	4,14.334	24. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral), AWS GovCloud (US-West), AWS GovCloud (US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 218,0	4,14.330	2. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral), AWS GovCloud (US-West), AWS GovCloud (US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202306,0	4,14.330	22. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada AWS GovCloud (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 116,0	4,14.328	11. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada AWS GovCloud (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023101,0	4,14.327	16. November 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada AWS GovCloud (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023020,1	4,14.326	7. November 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada AWS GovCloud (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023012,1	4,14.326	26. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada AWS GovCloud (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 926,0	4,14.322	19. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada AWS GovCloud (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)
2.0.2023 906,0	4,14.322	04. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 822.0	4,14.322	30. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.20230808,0	4,14.320	24. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 727.0	4,14.320	14. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 719,0	4,14.320	02. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 628.0	4,14.318	12. Juli 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 612,0	4,14.314	23. Juni 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 504.1	4,14.313	16. Mai 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 418,0	4,14.311	3. Mai 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 404.1	4,14.311	18. April 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)
2.0.2023 404.0	4,14.311	10. April 2023	USA Ost (Nord-Virginia), Europa (Paris)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 320.0	4,14.309	30. März 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202307,0	4,14.305	15. März 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202307,0	4,14.304	22. Februar 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023119.1	4,14.301	3. Februar 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022 210.1	4.14.301	22. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022103.3	4,14.296	5. Dezember 2022	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022004,0	4,14.294	02. November 2022	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022 912.1	4,14.291	7. Oktober 2022	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)
2.0.2022 719.0	4,14.287	10. August 2022	us-west-1 , eu-west-3 , eu-north-1 , ap-south-1 , me-south-1

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022.606.1	4.14.281	15. Juli 2022	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

## Bekannte Probleme

- Wenn Amazon-EMR-Version 6.5.0, 6.6.0 oder 6.7.0 Apache Phoenix-Tabellen über die Apache Spark-Shell liest, tritt `NoSuchMethodError` ein, weil Amazon EMR ein falsches `Hbase.compat.version` verwendet. Amazon EMR Version 6.8.0 behebt dieses Problem.
- Wenn Sie den DynamoDB-Konnektor mit Spark auf den Amazon-EMR-Versionen 6.6.0, 6.7.0 und 6.8.0 verwenden, geben alle Lesevorgänge aus Ihrer Tabelle ein leeres Ergebnis zurück, obwohl

der Eingabe-Split auf nicht leere Daten verweist. Das liegt daran, dass Spark 3.2.0 standardmäßig `spark.hadoopRDD.ignoreEmptySplits` auf `true` einstellt. Um das Problem zu umgehen, setzen Sie explizit `spark.hadoopRDD.ignoreEmptySplits` auf `false`. Amazon EMR Version 6.9.0 behebt dieses Problem.

- Wenn Sie Spark mit der Formatierung des Hive-Partitionsstandorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den Amazon-EMR-Versionen 5.30.0 bis 5.36.0 und 6.2.0 bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:
  - Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.
  - Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist `s3://bucket/table/p=a` ein Präfix von `s3://bucket/table/p=a b`.
  - Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen UTF-8-Wert, der kleiner als das `/`-Zeichen (U+002F) ist. Beispielsweise fällt das Leerzeichen (U+0020), das in `s3://bucket/table/p=a b` zwischen `a` und `b` vorkommt, in diese Kategorie. Beachten Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: `!"#$%&'()*+,-.` Weitere Informationen finden Sie unter [UTF-8-Kodierungstabelle und Unicode-Zeichen](#).

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.

- Bei den Amazon-EMR-Versionen 5.36.0 und 6.6.0 bis 6.9.0 kann es bei `SecretAgent`- und `RecordServer`-Servicekomponenten aufgrund einer falschen Konfiguration des Dateinamenmusters in den `Log4j2`-Eigenschaften zu einem Verlust von Protokolldaten kommen. Die falsche Konfiguration führt dazu, dass die Komponenten nur eine Protokolldatei pro Tag generieren. Wenn die Rotationsstrategie angewendet wird, überschreibt sie die vorhandene Datei, anstatt wie erwartet eine neue Protokolldatei zu generieren. Um das Problem zu umgehen, verwenden Sie eine Bootstrap-Aktion, um jede Stunde Protokolldateien zu generieren, und fügen Sie eine automatisch inkrementierte Ganzzahl an den Dateinamen an, um die Rotation zu handhaben.

Verwenden Sie für die Versionen Amazon EMR 6.6.0 bis 6.9.0 die folgende Bootstrap-Aktion, wenn Sie einen Cluster starten.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-6x/replace-puppet.sh,Args=[]"
```

Verwenden Sie für Amazon EMR 5.36.0 die folgende Bootstrap-Aktion, wenn Sie einen Cluster starten.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-5x/replace-puppet.sh,Args=[]"
```

- Die `GetClusterSessionCredentials`-API wird nicht mit Clustern unterstützt, die in Amazon EMR 6.7 oder niedriger ausgeführt werden.

## Version 6.6.0

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 6.6.0. Änderungen beziehen sich auf Version 6.5.0.

Erste Version: 09. Mai 2022

Letzte Aktualisierung der Dokumentation: 15. Juni 2022

### Neue Features

- Amazon EMR 6.6 unterstützt jetzt Apache Spark 3.2, Apache Spark RAPIDS 22.02, CUDA 11, Apache Hudi 0.10.1, Apache Iceberg 0.13, Trino 0.367 und PrestoDB 0.267.
- Wenn Sie einen Cluster mit der neuesten Patch-Version von Amazon EMR 5.36 oder höher, 6.6 oder höher oder 7.0 oder höher starten, verwendet Amazon EMR die neueste Version von Amazon Linux 2023 oder Amazon Linux 2 für das standardmäßige Amazon EMR-AMI. Weitere Informationen finden Sie unter [Verwenden des standardmäßigen Amazon-Linux-AMI für Amazon EMR](#).

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024 223.0	4,14.336	8. März 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten AWS GovCloud (Bahrain), Kanada (Zentral), AWS GovCloud (US-West), (US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024131,0	4,14.336	14. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral), AWS GovCloud (US-West), AWS GovCloud (US-Ost), China (Peking), China (Ningxia)



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024124,0	4,14.336	7. Februar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral), AWS GovCloud (US-West), AWS GovCloud (US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2024109,0	4,14.334	24. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral), AWS GovCloud (US-West), AWS GovCloud (US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 218,0	4,14.330	2. Januar 2024	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral), AWS GovCloud (US-West), AWS GovCloud (US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202306,0	4,14.330	22. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada AWS GovCloud (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 116,0	4,14.328	11. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada AWS GovCloud (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023101,0	4,14.327	16. November 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada AWS GovCloud (Zentral), AWS GovCloud (US-West),



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023020,1	4,14.326	7. November 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada AWS GovCloud (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023012,1	4,14.326	26. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada AWS GovCloud (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 926,0	4,14.322	19. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada AWS GovCloud (Zentral), AWS GovCloud (US-West),

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
			(US-Ost), China (Peking), China (Ningxia)
2.0.2023 906,0	4,14.322	04. Oktober 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 822.0	4,14.322	30. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral)



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.20230808.0	4,14.320	24. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 727.0	4,14.320	14. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Frankfurt), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Kanada (Zentral)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 719,0	4,14.320	02. August 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Europa (Stockholm), Europa (Mailand), Europa (Spanien), Europa (Frankfurt), Europa (Zürich), Europa (Irland), Europa (London), Europa (Paris), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Hyderabad), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Asien-Pazifik (Jakarta), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain), Naher Osten (VAE), Kanada (Zentral)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 628.0	4,14.318	12. Juli 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 612,0	4,14.314	23. Juni 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 504.1	4,14.313	16. Mai 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 418,0	4,14.311	3. Mai 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 404.1	4,14.311	18. April 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)
2.0.2023 404.0	4,14.311	10. April 2023	USA Ost (Nord-Virginia), Europa (Paris)



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023 320.0	4,14.309	30. März 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202307,0	4,14.305	15. März 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.202307,0	4,14.304	22. Februar 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2023119.1	4,14.301	3. Februar 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022 4.14.301 210.1		22. Dezember 2023	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022103.3	4,14.296	5. Dezember 2022	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022004,0	4,14.294	02. November 2022	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022 912.1	4,14.291	7. Oktober 2022	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)
2.0.2022 805.0	4,14.287	30. August 2022	us-west-1



OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022 719.0	4,14.287	10. August 2022	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022 426,0	4,14.281	10. Juni 2022	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

OsRelease Label (Amazon Linux-Version)	Amazon-Linux-Kernversion	Verfügbarkeitsdatum	Unterstützte Regionen
2.0.2022.406.1	4,14.275	2. Mai 2022	USA Ost (Nord-Virginia), USA Ost (Ohio), USA West (Nordkalifornien), USA West (Oregon), Kanada (Zentral), Europa (Stockholm), Europa (Irland), Europa (London), Europa (Paris), Europa (Frankfurt), Europa (Mailand), Asien-Pazifik (Hongkong), Asien-Pazifik (Mumbai), Asien-Pazifik (Jakarta), Asien-Pazifik (Tokio), Asien-Pazifik (Seoul), Asien-Pazifik (Osaka), Asien-Pazifik (Singapur), Asien-Pazifik (Sydney), Afrika (Kapstadt), Südamerika (São Paulo), Naher Osten (Bahrain)

- Mit Amazon EMR 6.6 und höher werden Anwendungen, die Log4j 1.x und Log4j 2.x verwenden, auf Log4j 1.2.17 (oder höher) bzw. Log4j 2.17.1 (oder höher) aktualisiert, sodass die [Bootstrap-Aktionen](#) die zur Behebung der CVE-Probleme bereitgestellt werden, nicht erforderlich sind.
- [Verwaltete Skalierung] Spark Shuffle Data Managed Scaling Optimization – Für Amazon-EMR-Versionen 5.34.0 und höher sowie EMR-Versionen 6.4.0 und höher unterstützt verwaltete Skalierung jetzt Spark Shuffle Data (Daten, die Spark partitionsübergreifend verteilt, um bestimmte Operationen auszuführen). Weitere Informationen zu Shuffle-Vorgängen finden Sie unter [Nutzung](#)

[von Managed Scaling in Amazon EMR](#) im Verwaltungshandbuch für Amazon EMR und [Spark-Programmierungs-Handbuch](#).

- Ab Amazon EMR 5.32.0 und 6.5.0 ist die dynamische Executor-Größenanpassung für Apache Spark standardmäßig aktiviert. Sie können dieses Feature ein- oder ausschalten, indem Sie den Konfigurationsparameter `spark.yarn.heterogeneousExecutors.enabled` verwenden.

### Änderungen, Verbesserungen und behobene Probleme

- Amazon EMR reduziert die Cluster-Startup-Zeit für Cluster, die die EMR-Standard-AMI-Option verwenden und nur gängige Anwendungen wie Apache Hadoop, Apache Spark und Apache Hive installieren, um durchschnittlich bis zu 80 Sekunden.

### Bekannte Probleme

- Wenn Amazon-EMR-Version 6.5.0, 6.6.0 oder 6.7.0 Apache Phoenix-Tabellen über die Apache Spark-Shell liest, tritt `NoSuchMethodError` ein, weil Amazon EMR ein falsches `Hbase.compat.version` verwendet. Amazon EMR Version 6.8.0 behebt dieses Problem.
- Wenn Sie den DynamoDB-Konnektor mit Spark auf den Amazon-EMR-Versionen 6.6.0, 6.7.0 und 6.8.0 verwenden, geben alle Lesevorgänge aus Ihrer Tabelle ein leeres Ergebnis zurück, obwohl der Eingabe-Split auf nicht leere Daten verweist. Das liegt daran, dass Spark 3.2.0 standardmäßig `spark.hadoopRDD.ignoreEmptySplits` auf `true` einstellt. Um das Problem zu umgehen, setzen Sie explizit `spark.hadoopRDD.ignoreEmptySplits` auf `false`. Amazon EMR Version 6.9.0 behebt dieses Problem.
- Auf Trino-Clustern mit langer Laufzeit aktiviert Amazon EMR 6.6.0 die Garbage-Collection-Protokollierungs-Parameter in der Trino `jvm.config`, um bessere Einblicke in die Garbage-Collection-Protokolle zu erhalten. Durch diese Änderung werden viele Garbage-Collection-Protokolle an die Datei `launcher.log` (`/var/log/trino/launcher.log`) angehängt. Wenn Sie Trino-Cluster in Amazon EMR 6.6.0 ausführen, kann es vorkommen, dass Knoten aufgrund der angehängten Protokolle nicht mehr genügend Festplattenspeicher haben, nachdem der Cluster einige Tage lang ausgeführt wurde.

Um dieses Problem zu umgehen, führen Sie das folgende Skript als Bootstrap-Aktion aus, um die Garbage-Collection-Protokollierungs-Parameter in `jvm.config` zu deaktivieren, während Sie den Cluster für Amazon EMR 6.6.0 erstellen oder klonen.

```
#!/bin/bash
```

```
set -ex
PRESTO_PUPPET_DIR='/var/aws/emr/bigtop-deploy/puppet/modules/trino'
sudo bash -c "sed -i '/-Xlog/d' ${PRESTO_PUPPET_DIR}/templates/jvm.config"
```

- Wenn Sie Spark mit der Formatierung des Hive-Partitionsstandorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den Amazon-EMR-Versionen 5.30.0 bis 5.36.0 und 6.2.0 bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:
  - Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.
  - Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist `s3://bucket/table/p=a` ein Präfix von `s3://bucket/table/p=a b`.
  - Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen UTF-8-Wert, der kleiner als das `/`-Zeichen (U+002F) ist. Beispielsweise fällt das Leerzeichen (U+0020), das in `s3://bucket/table/p=a b` zwischen `a` und `b` vorkommt, in diese Kategorie. Beachten Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: `!"#$%&'()*+,-.` Weitere Informationen finden Sie unter [UTF-8-Kodierungstabelle und Unicode-Zeichen](#).

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.

- Bei den Amazon-EMR-Versionen 5.36.0 und 6.6.0 bis 6.9.0 kann es bei `SecretAgent`- und `RecordServer`-Servicekomponenten aufgrund einer falschen Konfiguration des Dateinamenmusters in den `Log4j2`-Eigenschaften zu einem Verlust von Protokolldaten kommen. Die falsche Konfiguration führt dazu, dass die Komponenten nur eine Protokolldatei pro Tag generieren. Wenn die Rotationsstrategie angewendet wird, überschreibt sie die vorhandene Datei, anstatt wie erwartet eine neue Protokolldatei zu generieren. Um das Problem zu umgehen, verwenden Sie eine Bootstrap-Aktion, um jede Stunde Protokolldateien zu generieren, und fügen Sie eine automatisch inkrementierte Ganzzahl an den Dateinamen an, um die Rotation zu handhaben.

Verwenden Sie für die Versionen Amazon EMR 6.6.0 bis 6.9.0 die folgende Bootstrap-Aktion, wenn Sie einen Cluster starten.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-6x/replace-puppet.sh,Args=[]"
```

Verwenden Sie für Amazon EMR 5.36.0 die folgende Bootstrap-Aktion, wenn Sie einen Cluster starten.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-5x/replace-puppet.sh,Args=[]"
```

## Version 5.35.0

Dies ist der Versionshinweis zu Amazon-EMR-Version 5.35.0.

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.35.0. Änderungen beziehen sich auf Version 5.34.0.

Erste Version: 30. März 2022

### Neue Features

- Amazon-EMR-Anwendungen der Version 5.35, die Log4j 1.x und Log4j 2.x verwenden, werden auf Log4j 1.2.17 (oder höher) bzw. Log4j 2.17.1 (oder höher) aktualisiert und erfordern keine Bootstrap-Aktionen zur Minderung der CVE-Probleme in früheren Versionen. Siehe [Ansatz zur Minderung von CVE-2021-44228](#).

### Änderungen, Verbesserungen und behobene Probleme

#### Änderungen bei Flink

Änderungstyp	Beschreibung
Upgrades	<ul style="list-style-type: none"> <li>• Aktualisieren Sie die Flink-Version auf 1.14.2.</li> <li>• log4j wurde auf 2.17.1 aktualisiert.</li> </ul>

#### Hadoop-Änderungen

Änderungstyp	Beschreibung
Hadoop-Open-Source-Backports seit EMR 5.34.0	<ul style="list-style-type: none"> <li>• <a href="#">YARN-10438</a>: Behandle eine Container-ID mit Null in <code>clientRMService# () getContainerReport</code></li> </ul>

Änderungstyp	Beschreibung
	<ul style="list-style-type: none"><li>• <a href="#">YARN-7266: Threads für den Timeline-Server-Event-Handler sind gesperrt</a></li><li>• <a href="#">YARN-10438</a>: ATS 1.5 kann nicht gestartet werden, wenn Dateien beschädigt sind oder fehlen RollingLevelDb</li><li>• <a href="#">HADOOP-13500</a>: Synchronisieren der Iteration des Objekts mit den Konfigurationseigenschaften</li><li>• <a href="#">YARN-10651</a>: Ist bei eingeschaltetem NPE abgestürzt. CapacityScheduler AbstractYarnScheduler updateNodeResource()</li><li>• <a href="#">HDFS-12221</a>: Ersetzen Sie Xerces in XmlEditsVisitor</li><li>• <a href="#">HDFS-16410</a>: Unsicheres Einlesen von XML OfflineEditsXMLLoader</li></ul>
Hadoop-Änderungen und Korrekturen	<ul style="list-style-type: none"><li>• Tomcat, das in KMS und HttpFS verwendet wird, wurde auf 8.5.75 aktualisiert</li><li>• In FileSystemOptimizedCommitter Version 2 wurde die Erfolgsmarkierung in den CommitJob-Ausgabepfad geschrieben, der bei der Erstellung des Committers definiert wurde. Da sich die Ausgabepfade von CommitJob und Aufgabenebene unterscheiden können, wurde der Pfad korrigiert, sodass er den in den Manifestdateien definierten Pfad verwendet. Bei Hive-Aufträgen führt dies dazu, dass die Erfolgsmarkierung korrekt geschrieben wird, wenn Operationen wie dynamische Partition oder UNION ALL ausgeführt werden.</li></ul>

## Änderungen bei Hive

Änderungstyp	Beschreibung
<p>Hive wurde auf die <a href="#">Open-Source-Version 2.3.9</a> aktualisiert, einschließlich dieser JIRA-Korrekturen</p>	<ul style="list-style-type: none"> <li>• <a href="#">HIVE-17155</a>: findConfFile () in HiveConf .java hat einige Probleme mit dem Conf-Pfad</li> <li>• <a href="#">HIVE-24797</a>: Deaktiviert die Validierung von Standardwerten beim Parsen von Avro-Schemas</li> <li>• <a href="#">HIVE-21563</a>: Verbessern Sie die Leistung von Table#, indem Sie Once deaktivieren getEmptyTable registerAllFunctions</li> <li>• <a href="#">HIVE-18147</a>: Tests können mit java.net fehlschlagen. BindException: Adresse wird bereits verwendet</li> <li>• <a href="#">HIVE-24608</a>: Wechseln Sie im HMS-Client für Hive 2.3.x zurück zu get_table</li> <li>• <a href="#">HIVE-21200</a>: Vektorisierung — Die Datumsspalte wirft java.lang aus. UnsupportedOperationException für Parkett</li> <li>• <a href="#">HIVE-19228</a>: Entfernen Sie die Verwendung von commons-httpclient 3.x</li> </ul>
<p>Hive-Open-Source-Backports seit EMR 5.34.0</p>	<ul style="list-style-type: none"> <li>• <a href="#">HIVE-19990</a>: Eine Abfrage mit einem Intervall literal in der Join-Bedingung schlägt fehl</li> <li>• <a href="#">HIVE-25824</a>: Aktualisieren Sie Branch-2.3 auf log4j 2.17.0</li> <li>• <a href="#">TEZ-4062</a>: Die Planung spekulativer Versuche sollte abgebrochen werden, wenn die Aufgabe abgeschlossen ist</li> <li>• <a href="#">TEZ-4108</a>: Rennbedingungen NullPointerException während der spekulativen Ausführung</li> <li>• <a href="#">TEZ-3918</a>: Das Setzen von tez.task.log.level funktioniert nicht</li> </ul>



Änderungstyp	Beschreibung
Hive-Upgrades und -Korrekturen	<ul style="list-style-type: none"> <li>• Aktualisieren Sie die Log4j-Version auf 2.17.1</li> <li>• Auf ORC-Version 1.4.3 upgraden</li> <li>• Der Deadlock aufgrund eines fehlerhaften Thread-Ins wurde behoben ShuffleScheduler</li> </ul>
Neue Features	<ul style="list-style-type: none"> <li>• Feature zum Drucken von Hive-Abfragen in AM-Protokollen hinzugefügt. Diese ist standardmäßig deaktiviert. Flag/Konf.: <code>tez.am.emr.print.hive.query.in.log</code> Status (Standard): FALSE.</li> </ul>

### Oozie-Änderungen

Änderungstyp	Beschreibung
Oozie-Open-Source-Backports seit EMR 5.34.0	<ul style="list-style-type: none"> <li>• <a href="#">OOZIE-3652: Der Oozie-Launcher</a> sollte die Verzeichnisliste erneut versuchen, wenn dies der Fall ist NoSuchFileException</li> </ul>

### Pig-Änderungen

Änderungstyp	Beschreibung
Upgrades	<ul style="list-style-type: none"> <li>• log4j wurde auf 1.2.17 aktualisiert.</li> </ul>

### Bekannte Probleme

- Wenn Sie Spark mit der Formatierung des Hive-Partitionsstandorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den Amazon-EMR-Versionen 5.30.0 bis 5.36.0 und 6.2.0 bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:
  - Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.

- Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist `s3://bucket/table/p=a` ein Präfix von `s3://bucket/table/p=a b`.
- Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen UTF-8-Wert, der kleiner als das `/`-Zeichen (U+002F) ist. Beispielsweise fällt das Leerzeichen (U+0020), das in `s3://bucket/table/p=a b` zwischen `a` und `b` vorkommt, in diese Kategorie. Beachten Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: `!"#$%&'()*+,-.` Weitere Informationen finden Sie unter [UTF-8-Kodierungstabelle und Unicode-Zeichen](#).

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.

## Version 5.34.0

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.34.0. Änderungen beziehen sich auf Version 5.33.1.

Erste Version: 20. Januar 2022

Aktualisiertes Veröffentlichungsdatum: 21. März 2022

### Neue Features

- [Verwaltete Skalierung] Spark Shuffle Data Managed Scaling Optimization – Für Amazon-EMR-Versionen 5.34.0 und höher sowie EMR-Versionen 6.4.0 und höher unterstützt verwaltete Skalierung jetzt Spark Shuffle Data (Daten, die Spark partitionsübergreifend verteilt, um bestimmte Operationen auszuführen). Weitere Informationen zu Shuffle-Vorgängen finden Sie unter [Nutzung von Managed Scaling in Amazon EMR](#) im Verwaltungshandbuch für Amazon EMR und [Spark-Programmierungs-Handbuch](#).
- [Hudi] Verbesserungen zur Vereinfachung der Hudi-Konfiguration. Optimistische Parallelitätssteuerung standardmäßig deaktiviert.

### Änderungen, Verbesserungen und behobene Probleme

- Dies ist eine Version zur Behebung von Problemen mit Amazon-EMR-Skalierung, wenn ein Cluster nicht erfolgreich hoch-/herunterskaliert werden kann oder Anwendungsfehler verursacht werden.

- Zuvor führte ein manueller Neustart des Ressourcenmanagers auf einem Multi-Master-Cluster dazu, dass Amazon-EMR-Cluster-Daemons wie Zookeeper alle zuvor stillgelegten oder verlorenen Knoten in der Zookeeper znode-Datei neu geladen haben. Dies führte dazu, dass die Standardgrenzwerte in bestimmten Situationen überschritten wurden. Amazon EMR entfernt jetzt die außer Betrieb genommenen oder verlorenen Knotendatensätze, die älter als eine Stunde sind, aus der Zookeeper-Datei, und die internen Grenzwerte wurden erhöht.
- Es wurde ein Problem behoben, bei dem Skalierungsanforderungen für einen großen, stark ausgelasteten Cluster fehlschlagen, wenn Amazon-EMR-On-Cluster-Daemons Aktivitäten zur Integritätsprüfung durchführten, z. B. das Erfassen des YARN-Knotenstatus und des HDFS-Knotenstatus. Das lag daran, dass Cluster-Daemons die Integritätsstatusdaten eines Knotens nicht an interne Amazon-EMR-Komponenten weitergeben konnten.
- Verbesserte EMR-On-Cluster-Daemons zur korrekten Nachverfolgung der Knotenstatus bei der Wiederverwendung von IP-Adressen, um die Zuverlässigkeit bei Skalierungsvorgängen zu verbessern.
- [SPARK-29683](#). Es wurde ein Problem behoben, bei dem während der Cluster-Skalierung Auftragsfehler auftraten, da Spark davon ausging, dass alle verfügbaren Knoten auf der Verweigern-Liste standen.
- [YARN-9011](#). Es wurde ein Problem behoben, bei dem Auftragsfehler aufgrund eines Fehlers bei der Außerbetriebnahme von YARN auftraten, wenn der Cluster versuchte, hoch- oder herunterzuskalieren.
- Das Problem mit Schritt- oder Auftragsfehlern bei der Cluster-Skalierung wurde behoben, indem sichergestellt wurde, dass die Knotenstatus zwischen den Amazon-EMR-On-Cluster-Daemons und YARN/HDFS immer konsistent sind.
- Es wurde ein Problem behoben, bei dem Clustervorgänge wie Herunterskalierung und Schrittübermittlung für Amazon-EMR-Cluster, die mit Kerberos-Authentifizierung aktiviert waren, fehlschlagen. Dies lag daran, dass der Amazon-EMR-On-Cluster-Daemon das Kerberos-Ticket nicht erneuert hat, das für die sichere Kommunikation mit HDFS/YARN erforderlich ist, das auf dem Primärknoten ausgeführt wird.
- Zeppelin auf Version 0.10.0 upgegradet.
- Livy Fix – auf 0.7.1 aktualisiert
- Spark-Leistungsverbesserung – Heterogene Executors werden deaktiviert, wenn bestimmte Spark-Konfigurationswerte in EMR 5.34.0 überschrieben werden.

- WebHDFS- und HttpFS-Server werden standardmäßig deaktiviert. Sie können WebHDFS mithilfe der Hadoop-Konfiguration `dfs.webhdfs.enabled` erneut aktivieren. Der HttpFS-Server kann mit `sudo systemctl start hadoop-httpfs` gestartet werden.

## Bekannte Probleme

- Das Amazon-EMR-Notebooks-Feature, das mit dem Livy-Benutzerwechsel verwendet wird, funktioniert nicht, da HttpFS standardmäßig deaktiviert ist. In diesem Fall kann das EMR Notebook keine Verbindung zu dem Cluster herstellen, für den Livy-Identitätswechsel aktiviert ist. Die Problemlösung besteht darin, den HttpFS-Server zu starten, bevor Sie das EMR-Notebook mit dem Cluster `sudo systemctl start hadoop-httpfs` verbinden.
- Hue-Abfragen funktionieren in Amazon EMR 6.4.0 nicht, da der Apache-Hadoop-HttpFS-Server standardmäßig deaktiviert ist. Um Hue auf Amazon EMR 6.4.0 zu verwenden, starten Sie den HttpFS-Server auf dem Amazon-EMR-Primärknoten entweder manuell mithilfe von `sudo systemctl start hadoop-httpfs` oder [verwenden Sie einen Amazon-EMR-Schritt](#).
- Das Amazon-EMR-Notebooks-Feature, das mit dem Livy-Benutzerwechsel verwendet wird, funktioniert nicht, da HttpFS standardmäßig deaktiviert ist. In diesem Fall kann das EMR Notebook keine Verbindung zu dem Cluster herstellen, für den Livy-Identitätswechsel aktiviert ist. Die Problemlösung besteht darin, den HttpFS-Server zu starten, bevor Sie das EMR-Notebook mit dem Cluster `sudo systemctl start hadoop-httpfs` verbinden.
- Wenn Sie Spark mit der Formatierung des Hive-Partitionsstandorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den Amazon-EMR-Versionen 5.30.0 bis 5.36.0 und 6.2.0 bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:
  - Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.
  - Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist `s3://bucket/table/p=a` ein Präfix von `s3://bucket/table/p=a b`.
  - Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen UTF-8-Wert, der kleiner als das `/`-Zeichen (U+002F) ist. Beispielsweise fällt das Leerzeichen (U+0020), das in `s3://bucket/table/p=a b` zwischen `a` und `b` vorkommt, in diese Kategorie. Beachten Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: `!"#$%&'()*+,-.` Weitere Informationen finden Sie unter [UTF-8-Kodierungstabelle und Unicode-Zeichen](#).

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.

## Version 6.5.0

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 6.5.0. Änderungen beziehen sich auf Version 6.4.0.

Erste Version: 20. Januar 2022

Aktualisiertes Veröffentlichungsdatum: 21. März 2022

### Neue Features

- [Verwaltete Skalierung] Spark Shuffle Data Managed Scaling Optimization – Für Amazon-EMR-Versionen 5.34.0 und höher sowie EMR-Versionen 6.4.0 und höher unterstützt verwaltete Skalierung jetzt Spark Shuffle Data (Daten, die Spark partitionsübergreifend verteilt, um bestimmte Operationen auszuführen). Weitere Informationen zu Shuffle-Vorgängen finden Sie unter [Nutzung von Managed Scaling in Amazon EMR](#) im Verwaltungshandbuch für Amazon EMR und [Spark-Programmierungs-Handbuch](#).
- Ab Amazon EMR 5.32.0 und 6.5.0 ist die dynamische Executor-Größenanpassung für Apache Spark standardmäßig aktiviert. Sie können dieses Feature ein- oder ausschalten, indem Sie den Konfigurationsparameter `spark.yarn.heterogeneousExecutors.enabled` verwenden.
- Support für das offene Tabellenformat von Apache Iceberg für riesige analytische Datensätze.
- Support für ranger-trino-plugin 2.0.1-amzn-1
- Support für toree 0.5.0

### Änderungen, Verbesserungen und behobene Probleme

- Die Release-Version von Amazon EMR 6.5 unterstützt jetzt Apache Iceberg 0.12.0 und bietet Laufzeitverbesserungen mit Amazon-EMR-Laufzeit für Apache Spark, Amazon-EMR-Laufzeit für Presto und Amazon-EMR-Laufzeit für Apache Hive.
- [Apache Iceberg](#) ist ein offenes Tabellenformat für große Datensätze in Amazon S3 und bietet schnelle Abfrageleistung bei großen Tabellen, atomare Commits, gleichzeitige Schreibvorgänge

und eine SQL-kompatible Tabellenentwicklung. Mit EMR 6.5 können Sie Apache Spark 3.1.2 mit dem Iceberg-Tabellenformat verwenden.

- Apache Hudi 0.9 fügt Spark SQL DDL- und DML-Unterstützung hinzu. Auf diese Weise können Sie Hudi-Tabellen nur mit SQL-Anweisungen erstellen und ändern. Apache Hudi 0.9 beinhaltet auch Leistungsverbesserungen auf der Abfrageseite und auf der Writer-Seite.
- Amazon-EMR-Laufzeit für Apache Hive verbessert die Leistung von Apache Hive auf Amazon S3, indem Umbenennungsvorgänge bei Staging-Vorgängen entfernt werden, und verbessert die Leistung von Metastore Check (MSCK)-Befehlen, die zur Reparatur von Tabellen verwendet werden.

## Bekannte Probleme

- Wenn Amazon-EMR-Version 6.5.0, 6.6.0 oder 6.7.0 Apache Phoenix-Tabellen über die Apache Spark-Shell liest, tritt `NoSuchMethodError` ein, weil Amazon EMR ein falsches `Hbase.compat.version` verwendet. Amazon EMR Version 6.8.0 behebt dieses Problem.
- Hbase-Bundle-Cluster mit hoher Verfügbarkeit (HA) können nicht mit der Standard-Volumen-Größe und dem Instance-Typ bereitgestellt werden. Dieses Problem lässt sich umgehen, indem Sie die Größe des Root-Volumens erhöhen.
- Um Spark-Aktionen mit Apache Oozie zu verwenden, müssen Sie Ihrer `Oozie-workflow.xml`-Datei die folgende Konfiguration hinzufügen. Andernfalls fehlen mehrere wichtige Bibliotheken wie Hadoop und EMRFS im Klassenpfad der Spark-Executoren, die Oozie startet.

```
<spark-opts>--conf spark.yarn.populateHadoopClasspath=true</spark-opts>
```

- Wenn Sie Spark mit der Formatierung des Hive-Partitionsstandorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den Amazon-EMR-Versionen 5.30.0 bis 5.36.0 und 6.2.0 bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:
  - Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.
  - Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist `s3://bucket/table/p=a` ein Präfix von `s3://bucket/table/p=a b`.
  - Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen UTF-8-Wert, der kleiner als das `/`-Zeichen (U+002F) ist. Beispielsweise fällt das Leerzeichen (U+0020), das in `s3://bucket/table/p=a b` zwischen `a` und `b` vorkommt, in diese Kategorie. Beachten

Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: !"#%&'()\*+,-. Weitere Informationen finden Sie unter [UTF-8-Kodierungstabelle und](#) Unicode-Zeichen.

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.

## Version 6.4.0

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 6.4.0. Änderungen beziehen sich auf Version 6.3.0.

Erste Version: 20. September 2021

Aktualisiertes Veröffentlichungsdatum: 21. März 2022

### Unterstützte Anwendungen

- AWS SDK for Java Ausführung 1.12.31
- CloudWatch Version 2.2.0 senken
- Version 4.16.0 von EMR DynamoDB Connector
- EMRFS-Version 2.47.0
- Version 3.2.0 von Amazon EMR Goodies
- Version 3.5.0 von Amazon EMR Kinesis Connector
- Amazon-EMR-Aufzeichnungsserver Version 2.1.0
- Version 2.5.0 von Amazon EMR Skripts
- Flink, Version 1.13.1
- Ganglia Version 3.7.2
- AWS Glue Hive Metastore Client Version 3.3.0
- Hadoop Version 3.2.1-amzn-4
- HBase Version 2.4.4-amzn-0
- ase-operator-tools HB 1.1.0
- HCatalog Version 3.1.2-amzn-5
- Hive Version 3.1.2-amzn-5
- Hudi Version 0.8.0-amzn-0

- Hue Version 4.9.0
- Java JDK-Version Corretto-8.302.08.1 (Build 1.8.0\_302-b08)
- JupyterHub Ausführung 1.4.1
- Livy Version 0.7.1-incubating
- MXNet-Version 1.8.0
- Oozie-Version 5.2.1
- Phoenix Version 5.1.2
- Pig Version 0.17.0
- Presto Version 0.254.1-amzn-0
- Trino Version 359
- Version 2.0.0 von Apache Ranger KMS (transparente Multi-Master-Verschlüsselung)
- Ranger-Plugins 2.0.1-amzn-0
- Ranger-S3-Plugin 1.2.0
- SageMaker Spark SDK, Version 1.4.1
- Scala-Version 2.12.10 (OpenJDK 64-Bit-Server-VM, Java 1.8.0\_282)
- Spark Version 3.1.2-amzn-0
- Spark-Rapids 0.4.1
- Sqoop-Version 1.4.7
- TensorFlow Ausführung 2.4.1
- Tez-Version 0.9.2
- Zeppelin-Version 0.9.0
- Zookeeper Version 3.5.7
- Konnektoren und Treiber: DynamoDB Connector 4.16.0

## Neue Features

- [Verwaltete Skalierung] Spark Shuffle Data Managed Scaling Optimization – Für Amazon-EMR-Versionen 5.34.0 und höher sowie EMR-Versionen 6.4.0 und höher unterstützt verwaltete Skalierung jetzt Spark Shuffle Data (Daten, die Spark partitionsübergreifend verteilt, um bestimmte Operationen auszuführen). Weitere Informationen zu Shuffle-Vorgängen finden Sie unter [Nutzung von Managed Scaling in Amazon EMR](#) im Verwaltungshandbuch für Amazon EMR und [Spark-Programmierungs-Handbuch](#).



- Auf Apache-Ranger-fähigen Amazon-EMR-Clustern können Sie Apache Spark SQL verwenden, um Daten mit, INSERT INTO, INSERT OVERWRITE und ALTER TABLE in die Apache-Hive-Metastore-Tabellen einzufügen oder diese zu aktualisieren. Wenn Sie ALTER TABLE mit Spark SQL verwenden, muss ein Partitionsspeicherort das untergeordnete Verzeichnis eines Tabellenspeicherorts sein. Amazon EMR unterstützt derzeit nicht das Einfügen von Daten in eine Partition, deren Partitionsspeicherort sich von der Tabellenposition unterscheidet.
- PrestoSQL wurde [in Trino umbenannt](#).
- Hive: Die Ausführung einfacher SELECT-Abfragen mit LIMIT-Klausel wird beschleunigt, indem die Abfrageausführung gestoppt wird, sobald die in der LIMIT-Klausel angegebene Anzahl von Datensätzen abgerufen wurde. Einfache SELECT-Abfragen sind Abfragen ohne GROUP BY/ORDER BY-Klausel oder Abfragen ohne Reducer-Stage. z. B. SELECT \* from <TABLE> WHERE <Condition> LIMIT <Number>.

## Hudi-Parallelitätskontrolle

- Hudi unterstützt jetzt Optimistic Concurrency Control (OCC – Optimistische Parallelitätskontrolle), die mit Schreiboperationen wie UPSERT und INSERT genutzt werden kann, um Änderungen von mehreren Autoren an derselben Hudi-Tabelle zu ermöglichen. Dies ist OCC auf Dateiebene, sodass zwei beliebige Commits (oder Writer) in dieselbe Tabelle schreiben können, sofern ihre Änderungen nicht kollidieren. Weitere Informationen finden Sie unter [Hudi-Parallelitätskontrolle..](#)
- Auf Amazon-EMR-Clustern ist Zookeeper installiert, der als Sperranbieter für OCC genutzt werden kann. Um die Verwendung dieses Features zu vereinfachen, sind in Amazon-EMR-Clustern die folgenden Eigenschaften vorkonfiguriert:

```
hoodie.write.lock.provider=org.apache.hudi.client.transaction.lock.ZookeeperBasedLockProvider
hoodie.write.lock.zookeeper.url=<EMR Zookeeper URL>
hoodie.write.lock.zookeeper.port=<EMR Zookeeper Port>
hoodie.write.lock.zookeeper.base_path=/hudi
```

Um OCC zu aktivieren, müssen Sie die folgenden Eigenschaften entweder mit ihren Hudi-Auftragsoptionen oder auf Cluster-Ebene mithilfe der Amazon-EMR-Konfigurations-API konfigurieren:

```
hoodie.write.concurrency.mode=optimistic_concurrency_control
hoodie.cleaner.policy.failed.writes=LAZY (Performs cleaning of failed writes lazily instead of inline with every write)
```

```
hoodie.write.lock.zookeeper.lock_key=<Key to uniquely identify the Hudi table> (Table Name is a good option)
```

## Hudi Monitoring: CloudWatch Amazon-Integration zur Berichterstattung über Hudi-Metriken

- Amazon EMR unterstützt die Veröffentlichung von Hudi Metrics auf Amazon. CloudWatch Es wird aktiviert, indem die folgenden erforderlichen Konfigurationen festgelegt werden:

```
hoodie.metrics.on=true
hoodie.metrics.reporter.type=CLOUDWATCH
```

- Im Folgenden finden Sie optionale Hudi-Konfigurationen, die Sie ändern können:

Einstellung	Beschreibung	Wert
hoodie.metrics.cloudwatch.report.period.seconds	Häufigkeit (in Sekunden), mit der Kennzahlen an Amazon gemeldet werden CloudWatch	Der Standardwert ist 60s, was für die von Amazon angebotene Standardauflösung von einer Minute in Ordnung ist. CloudWatch
hoodie.metrics.cloudwatch.metric.prefix	Präfix, das jedem Metriknamen hinzugefügt werden soll	Der Standardwert ist leer (kein Präfix)
hoodie.metrics.cloudwatch.namespace	CloudWatch Amazon-Namespace, unter dem Metriken veröffentlicht werden	Der Standardwert ist Hudi
hoodie.metrics.cloudwatch.maxDatumsPerAnfrage	Maximale Anzahl von Daten, die in einer Anfrage an Amazon enthalten sein können CloudWatch	Der Standardwert ist 20, was dem CloudWatch Amazon-Standard entspricht

## Unterstützung und Verbesserungen für Amazon-EMR-Hudi-Konfigurationen

- Kunden können jetzt die EMR-Konfigurations-API und das Rekonfigurationsfeature nutzen, um Hudi-Konfigurationen auf Cluster-Ebene zu konfigurieren. Eine neue dateibasierte

Konfigurationsunterstützung wurde über `/etc/hudi/conf/hudi-defaults.conf` eingeführt, ähnlich wie bei anderen Anwendungen wie Spark, Hive usw. EMR konfiguriert einige Standardwerte, um die Benutzererfahrung zu verbessern:

- `hoodie.datasource.hive_sync.jdbcurl` ist für die Cluster-Hive-Server-URL konfiguriert und muss nicht mehr angegeben werden. Dies ist besonders nützlich, wenn Sie einen Auftrag im Spark-Cluster-Modus ausführen, wo Sie zuvor die Amazon-EMR-Master-IP angeben mussten.
- HBase-spezifische Konfigurationen, die für die Verwendung des HBase-Index mit Hudi nützlich sind.
- Spezifische Konfiguration des Zookeeper-Sperranbieters, wie unter Parallelitätskontrolle beschrieben, was die Verwendung von Optimistischer Parallelitätskontrolle (OCC) erleichtert.
- Zusätzliche Änderungen wurden eingeführt, um die Anzahl der Konfigurationen zu reduzieren, die Sie bestehen müssen, und um nach Möglichkeit automatische Schlüsse zu ziehen:
  - Das `partitionBy` -Schlüsselwort kann verwendet werden, um die Partitionsspalte zu spezifizieren.
  - Bei der Aktivierung von Hive Sync ist es nicht mehr erforderlich, den Vorgang `HIVE_TABLE_OPT_KEY`, `HIVE_PARTITION_FIELDS_OPT_KEY`, `HIVE_PARTITION_EXTRACTOR_CLASS_OPT_KEY` zu bestehen. Diese Werte können aus dem Hudi-Tabellennamen und dem Partitionsfeld abgeleitet werden.
  - `KEYGENERATOR_CLASS_OPT_KEY` ist nicht zwingend erforderlich und kann aus einfacheren Fällen von `SimpleKeyGenerator` und `ComplexKeyGenerator` abgeleitet werden.

## Vorbehalte von Hudi

- Hudi unterstützt keine vektorisierte Ausführung in Hive für Merge on Read (MoR)- und Bootstrap-Tabellen. So schlägt beispielsweise `count(*)` mit der Hudi-Echtzeittabelle fehl, wenn `hive.vectorized.execution.enabled` auf „wahr“ gesetzt ist. Um das Problem zu umgehen, können Sie das vektorisierte Lesen deaktivieren, indem Sie `hive.vectorized.execution.enabled` auf `false` setzen.
- Die Multi-Writer-Unterstützung ist nicht mit dem Hudi-Bootstrap-Feature kompatibel.
- Flink Streamer und Flink SQL sind experimentelle Features in dieser Version. Diese Features werden nicht zur Verwendung in Produktionsbereitstellungen empfohlen.

## Änderungen, Verbesserungen und behobene Probleme

Dies ist eine Version zur Behebung von Problemen mit Amazon-EMR-Skalierung, wenn ein Cluster nicht erfolgreich hoch-/herunterskaliert werden kann oder Anwendungsfehler verursacht werden.

- Zuvor führte ein manueller Neustart des Ressourcenmanagers auf einem Multi-Master-Cluster dazu, dass Amazon-EMR-Cluster-Daemons wie Zookeeper alle zuvor stillgelegten oder verlorenen Knoten in der Zookeeper znode-Datei neu geladen haben. Dies führte dazu, dass die Standardgrenzwerte in bestimmten Situationen überschritten wurden. Amazon EMR entfernt jetzt die außer Betrieb genommenen oder verlorenen Knotendatensätze, die älter als eine Stunde sind, aus der Zookeeper-Datei, und die internen Grenzwerte wurden erhöht.
- Es wurde ein Problem behoben, bei dem Skalierungsanforderungen für einen großen, stark ausgelasteten Cluster fehlschlagen, wenn Amazon-EMR-On-Cluster-Daemons Aktivitäten zur Integritätsprüfung durchführten, z. B. das Erfassen des YARN-Knotenstatus und des HDFS-Knotenstatus. Das lag daran, dass Cluster-Daemons die Integritätsstatusdaten eines Knotens nicht an interne Amazon-EMR-Komponenten weitergeben konnten.
- Verbesserte EMR-On-Cluster-Daemons zur korrekten Nachverfolgung der Knotenstatus bei der Wiederverwendung von IP-Adressen, um die Zuverlässigkeit bei Skalierungsvorgängen zu verbessern.
- [SPARK-29683](#). Es wurde ein Problem behoben, bei dem während der Cluster-Skalierung Auftragsfehler auftraten, da Spark davon ausging, dass alle verfügbaren Knoten auf der Verweigern-Liste standen.
- [YARN-9011](#). Es wurde ein Problem behoben, bei dem Auftragsfehler aufgrund eines Fehlers bei der Außerbetriebnahme von YARN auftraten, wenn der Cluster versuchte, hoch- oder herunterzuskalieren.
- Das Problem mit Schritt- oder Auftragsfehlern bei der Cluster-Skalierung wurde behoben, indem sichergestellt wurde, dass die Knotenstatus zwischen den Amazon-EMR-On-Cluster-Daemons und YARN/HDFS immer konsistent sind.
- Es wurde ein Problem behoben, bei dem Clustervorgänge wie Herunterskalierung und Schrittübermittlung für Amazon-EMR-Cluster, die mit Kerberos-Authentifizierung aktiviert waren, fehlschlagen. Dies lag daran, dass der Amazon-EMR-On-Cluster-Daemon das Kerberos-Ticket nicht erneuert hat, das für die sichere Kommunikation mit HDFS/YARN erforderlich ist, das auf dem Primärknoten ausgeführt wird.
- Konfigurieren eines Clusters zur Behebung von Leistungsproblemen mit Apache YARN Timeline Server Version 1 und 1.5

Apache YARN Timeline Server Version 1 und 1.5 können bei sehr aktiven, großen EMR-Clustern zu Leistungsproblemen führen, insbesondere bei `yarn.resourcemanager.system-metrics-publisher.enabled=true`, was die Standardeinstellung in Amazon EMR ist. Ein Open-Source-YARN-Timeline-Server v2 löst das Leistungsproblem im Zusammenhang mit der Skalierbarkeit des YARN Timeline Servers.

Weitere Lösungen für dieses Problem umfassen:

- Konfiguration von `yarn.resourcemanager.system-metrics-publisher.enabled=false` in `yarn-site.xml`.
- Aktivieren des Fixes für dieses Problem beim Erstellen eines Clusters, wie unten beschrieben.

Die folgenden Amazon-EMR-Versionen enthalten eine Lösung für dieses Leistungsproblem mit YARN Timeline Server.

EMR 5.30.2, 5.31.1, 5.32.1, 5.33.1, 5.34.x, 6.0.1, 6.1.1, 6.2.1, 6.3.1, 6.4.x

Um den Fix für eine der oben angegebenen Amazon-EMR-Versionen zu aktivieren, legen Sie diese Eigenschaften `true` in einer JSON-Konfigurationsdatei fest, die mit dem [aws emr create-cluster-Befehlsparameter](#) `--configurations file://./configurations.json` übergeben wird. Oder aktivieren Sie den Fix über die Benutzeroberfläche der [Rekonfigurationskonsole](#).

Beispiel für den Inhalt der Datei `configurations.json`:

```
[
{
  "Classification": "yarn-site",
  "Properties": {
    "yarn.resourcemanager.system-metrics-publisher.timeline-server-v1.enable-batch":
      "true",
    "yarn.resourcemanager.system-metrics-publisher.enabled": "true"
  },
  "Configurations": []
}
```

- WebHDFS- und HttpFS-Server werden standardmäßig deaktiviert. Sie können WebHDFS mithilfe der Hadoop-Konfiguration `dfs.webhdfs.enabled` erneut aktivieren. Der HttpFS-Server kann mit `sudo systemctl start hadoop-httpfs` gestartet werden.

- HTTPS ist jetzt standardmäßig für Amazon-Linux-Repositorys aktiviert. Wenn Sie eine Amazon-S3-VPCE-Richtlinie verwenden, um den Zugriff auf bestimmte Buckets zu beschränken, müssen Sie den neuen Amazon-Linux-Bucket-ARN `arn:aws:s3:::amazonlinux-2-repos-$region/*` zu Ihrer Richtlinie hinzufügen (ersetzen Sie `$region` durch die Region, in der sich der Endpunkt befindet). Weitere Informationen finden Sie unter diesem Thema in den Diskussionsforen. [AWS Ankündigung: Amazon Linux 2 unterstützt jetzt die Möglichkeit, HTTPS zu verwenden, wenn eine Verbindung zu Paket-Repositorys hergestellt wird.](#)
- Hive: Die Leistung bei Schreibabfragen wurde verbessert, indem die Verwendung eines Scratch-Verzeichnisses auf HDFS für den letzten JAuftrag aktiviert wurde. Die temporären Daten für den endgültigen Auftrag werden in HDFS statt in Amazon S3 geschrieben, und die Leistung wurde verbessert, da die Daten von HDFS an den endgültigen Tabellenort (Amazon S3) und nicht zwischen Amazon-S3-Geräten verschoben werden.
- Hive: Verbesserung der Kompilierungszeit von Abfragen um das 2,5-fache mit Glue Metastore Partition Pruning.
- Wenn integrierte UDFs von Hive an den Hive-Metastore-Server übergeben werden, wird standardmäßig nur eine Teilmenge dieser integrierten UDFs an den Glue Metastore übergeben, da Glue nur begrenzte Ausdrucksoperatoren unterstützt. Wenn Sie `hive.glue.partition.pruning.client=true` festlegen, erfolgt das gesamte Partitionsbereinigen auf der Clientseite. Wenn Sie `hive.glue.partition.pruning.server=true` festlegen, erfolgt das gesamte Bereinigen von Partitionen auf der Serverseite.

## Bekannte Probleme

- Hue-Abfragen funktionieren in Amazon EMR 6.4.0 nicht, da der Apache-Hadoop-HttpFS-Server standardmäßig deaktiviert ist. Um Hue auf Amazon EMR 6.4.0 zu verwenden, starten Sie den HttpFS-Server auf dem Amazon-EMR-Primärknoten entweder manuell mithilfe von `sudo systemctl start hadoop-httpfs` oder [verwenden Sie einen Amazon-EMR-Schritt.](#)
- Das Amazon-EMR-Notebooks-Feature, das mit dem Livy-Benutzerwechsel verwendet wird, funktioniert nicht, da HttpFS standardmäßig deaktiviert ist. In diesem Fall kann das EMR Notebook keine Verbindung zu dem Cluster herstellen, für den Livy-Identitätswechsel aktiviert ist. Die Problemumgehung besteht darin, den HttpFS-Server zu starten, bevor Sie das EMR-Notebook mit dem Cluster `sudo systemctl start hadoop-httpfs` verbinden.
- In Amazon-EMR-Version 6.4.0 unterstützt Phoenix die Komponente Phoenix-Konnektoren nicht.

- Um Spark-Aktionen mit Apache Oozie zu verwenden, müssen Sie Ihrer Oozie-workflow.xml-Datei die folgende Konfiguration hinzufügen. Andernfalls fehlen mehrere wichtige Bibliotheken wie Hadoop und EMRFS im Klassenpfad der Spark-Executoren, die Oozie startet.

```
<spark-opts>--conf spark.yarn.populateHadoopClasspath=true</spark-opts>
```

- Wenn Sie Spark mit der Formatierung des Hive-Partitionsstandorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den Amazon-EMR-Versionen 5.30.0 bis 5.36.0 und 6.2.0 bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:
  - Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.
  - Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist `s3://bucket/table/p=a` ein Präfix von `s3://bucket/table/p=a b`.
  - Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen UTF-8-Wert, der kleiner als das `/`-Zeichen (U+002F) ist. Beispielsweise fällt das Leerzeichen (U+0020), das in `s3://bucket/table/p=a b` zwischen `a` und `b` vorkommt, in diese Kategorie. Beachten Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: `!"#$%&'()*+,-.` Weitere Informationen finden Sie unter [UTF-8-Kodierungstabelle und Unicode-Zeichen](#).

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.

## Version 5.32.0

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.32.0. Änderungen beziehen sich auf Version 5.31.0.

Erste Version: 8. Januar 2021

### Upgrades

- Der Amazon-Glue-Konnektor wurde auf Version 1.14.0 aktualisiert
- Amazon SageMaker Spark SDK wurde auf Version 1.4.1 aktualisiert
- Auf Version AWS SDK for Java 1.11.890 aktualisiert
- EMR DynamoDB Connector wurde auf Version 4.16.0 aktualisiert

- EMRFS auf Version 2.45.0 aktualisiert
- EMR-Log-Analytics-Metriken wurden auf Version 1.18.0 aktualisiert
- MetricsAndEventsApiGateway EMR-Client auf Version 1.5.0 aktualisiert
- EMR Record Server auf Version 1.8.0 aktualisiert
- EMR S3 Dist CP auf Version 2.17.0 aktualisiert
- EMR Record Server auf Version 1.7.0 aktualisiert
- Flink auf Version 1.11.2 aktualisiert
- Hadoop wurde auf Version 2.10.1-amzn-0 aktualisiert
- Upgrade von Hive auf Version 2.3.7-amzn-3
- Hue auf Version 4.8.0 aktualisiert
- Mxnet wurde auf Version 1.7.0 aktualisiert
- OpenCV wurde auf Version 4.4.0 aktualisiert
- Presto auf Version 0.240.1-amzn-0 aktualisiert
- Spark auf Version 2.4.7-amzn-0 aktualisiert
- Auf Version TensorFlow 2.3.1 aktualisiert

### Änderungen, Verbesserungen und behobene Probleme

- Dies ist eine Version zur Behebung von Problemen mit Amazon-EMR-Skalierung, wenn ein Cluster nicht erfolgreich hoch-/herunterskaliert werden kann oder Anwendungsfehler verursacht werden.
- Es wurde ein Problem behoben, bei dem Skalierungsanforderungen für einen großen, stark ausgelasteten Cluster fehlschlagen, wenn Amazon-EMR-On-Cluster-Daemons Aktivitäten zur Integritätsprüfung durchführten, z. B. das Erfassen des YARN-Knotenstatus und des HDFS-Knotenstatus. Das lag daran, dass Cluster-Daemons die Integritätsstatusdaten eines Knotens nicht an interne Amazon-EMR-Komponenten weitergeben konnten.
- Verbesserte EMR-On-Cluster-Daemons zur korrekten Nachverfolgung der Knotenstatus bei der Wiederverwendung von IP-Adressen, um die Zuverlässigkeit bei Skalierungsvorgängen zu verbessern.
- [SPARK-29683](#). Es wurde ein Problem behoben, bei dem während der Cluster-Skalierung Auftragsfehler auftraten, da Spark davon ausging, dass alle verfügbaren Knoten auf der Verweigern-Liste standen.



- [YARN-9011](#). Es wurde ein Problem behoben, bei dem Auftragsfehler aufgrund eines Fehlers bei der Außerbetriebnahme von YARN auftraten, wenn der Cluster versuchte, hoch- oder herunterzuskalieren.
- Das Problem mit Schritt- oder Auftragsfehlern bei der Cluster-Skalierung wurde behoben, indem sichergestellt wurde, dass die Knotenstatus zwischen den Amazon-EMR-On-Cluster-Daemons und YARN/HDFS immer konsistent sind.
- Es wurde ein Problem behoben, bei dem Clustervorgänge wie Herunterskalierung und Schrittübermittlung für Amazon-EMR-Cluster, die mit Kerberos-Authentifizierung aktiviert waren, fehlschlugen. Dies lag daran, dass der Amazon-EMR-On-Cluster-Daemon das Kerberos-Ticket nicht erneuert hat, das für die sichere Kommunikation mit HDFS/YARN erforderlich ist, das auf dem Primärknoten ausgeführt wird.
- Neuere Amazon-EMR-Versionen beheben das Problem mit einem niedrigeren Limit für die „maximale Anzahl geöffneter Dateien“ auf älteren AL2-Versionen in Amazon EMR. Die Amazon-EMR-Versionen 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 und höher enthalten jetzt einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“.
- Aktualisierte Komponentenversionen.
- Eine Liste der Komponentenversionen finden Sie in diesem Handbuch unter [Über Amazon-EMR-Versionen](#).

## Neue Features

- Ab Amazon EMR 5.32.0 und 6.5.0 ist die dynamische Executor-Größenanpassung für Apache Spark standardmäßig aktiviert. Sie können dieses Feature ein- oder ausschalten, indem Sie den Konfigurationsparameter `spark.yarn.heterogeneousExecutors.enabled` verwenden.
- Supportstatus für Instance Metadata Service (IMDS) V2: Amazon-EMR-Komponenten von 5.23.1, 5.27.1 und 5.32 oder höher verwenden IMDSv2 für alle IMDS-Aufrufe. Für IMDS-Aufrufe in Ihrem Anwendungscode können Sie sowohl IMDSv1 als auch IMDSv2 verwenden oder das IMDS so konfigurieren, dass es aus Sicherheitsgründen nur IMDSv2 verwendet. Bei anderen 5.x-EMR-Versionen führt die Deaktivierung von IMDSv1 zu einem Cluster-Startup-Fehler.
- Mit Amazon EMR 5.32.0 können Sie einen Cluster starten, der nativ in Apache Ranger integriert ist. Apache Ranger ist ein Open-Source-Framework zur Aktivierung, Überwachung und Verwaltung einer umfassenden Datensicherheit auf der gesamten Hadoop-Plattform. Weitere Informationen finden Sie unter [Apache Ranger](#). Dank der nativen Integration können Sie Ihren eigenen Apache Ranger verwenden, um eine detaillierte Datenzugriffskontrolle auf Amazon EMR durchzusetzen.

Weitere Informationen finden Sie unter [Amazon EMR mit Apache Ranger](#) im Verwaltungshandbuch für Amazon EMR.

- Amazon-EMR-Version 5.32.0 unterstützt Amazon EMR in EKS. Weitere Informationen zu den ersten Schritten mit EMR auf EKS finden Sie unter [Was ist Amazon EMR in EKS](#).
- Amazon-EMR-Version 5.32.0 unterstützt Amazon EMR Studio (Vorversion). Weitere Informationen zu den ersten Schritten mit EMR Studio finden Sie unter [Amazon EMR Studio \(Vorversion\)](#).
- **Verwaltete Richtlinien mit Geltungsbereich:** Um den AWS bewährten Methoden zu entsprechen, hat Amazon EMR verwaltete Standardrichtlinien mit EMR-Geltungsbereich der Version v2 eingeführt, die als Ersatz für Richtlinien dienen, die nicht mehr unterstützt werden. Weitere Informationen finden Sie unter [Verwaltete Richtlinien von Amazon EMR](#).

## Bekannte Probleme

- Für private Subnetz-Cluster von Amazon EMR 6.3.0 und 6.2.0 können Sie nicht auf die Ganglia-Webbenutzeroberfläche zugreifen. Sie erhalten die Fehlermeldung „Zugriff verweigert (403)“. Andere Web-Benutzeroberflächen wie Spark, Hue, Zeppelin, Livy und Tez funktionieren JupyterHub normal. Der Zugriff auf die Ganglia-Web-Benutzeroberfläche auf öffentlichen Subnetzclustern funktioniert ebenfalls normal. Um dieses Problem zu beheben, starten Sie den httpd-Service auf dem Primärknoten mit `sudo systemctl restart httpd` neu. Dieses Problem wurde in Amazon EMR 6.4.0 behoben.
- Niedrigeres Limit für die „Max. Anzahl geöffneter Dateien“ auf älteren AL2-Versionen [in neueren Versionen behoben]. Amazon-EMR-Versionen: emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0 und emr-6.2.0 basieren auf älteren Versionen von Amazon Linux 2 (AL2), die eine niedrigere Ulimit-Einstellung für „Max. Anzahl geöffneter Dateien“ haben, wenn Amazon-EMR-Cluster mit dem Standard-AMI erstellt werden. Die Amazon-EMR-Versionen 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 und höher enthalten einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“. Versionen mit einem niedrigeren Limit für geöffnete Dateien verursachen beim Senden des Spark-Jobs den Fehler „Zu viele offene Dateien“. In den betroffenen Versionen hat das Amazon-EMR-Standard-AMI eine Ulimit-StandardEinstellung von 4 096 für „Max. Anzahl geöffneter Dateien“, was unter dem Dateilimit von 65 536 im neuesten Amazon-Linux-2-AMI liegt. Die niedrigere Ulimit-Einstellung für „Max. Anzahl geöffneter Dateien“ führt dazu, dass der Spark-Job fehlschlägt, wenn der Spark-Treiber und der Executor versuchen, mehr als 4 096 Dateien zu öffnen. Um das Problem zu beheben, verfügt Amazon EMR über ein Bootstrap Action (BA)-Skript, das die Ulimit-Einstellung bei der Cluster-Erstellung anpasst.

Wenn Sie eine ältere Amazon-EMR-Version verwenden, für die dieses Problem nicht dauerhaft behoben ist, können Sie mit der folgenden Problemumgehung das Instance-Controller-Ulimit explizit auf maximal 65 536 Dateien festlegen.

Explizit ein ulimit über die Befehlszeile setzen

1. Bearbeiten Sie `/etc/systemd/system/instance-controller.service`, um die folgenden Parameter zum Abschnitt Service hinzuzufügen.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Starten Sie neu InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Mithilfe der Bootstrap-Aktion (BA) ein Ulimit festlegen

Sie können auch ein Bootstrap-Aktionsskript (BA) verwenden, um das Ulimit für den Instance-Controller bei der Clustererstellung auf 65 536 Dateien zu konfigurieren.

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
```

```
sudo systemctl daemon-reload
```

**⚠ Important**

EMR-Cluster, auf denen Amazon-Linux- oder Amazon-Linux-2-AMIs (Amazon Machine Images) ausgeführt werden, verwenden das Standardverhalten von Amazon Linux und laden wichtige und kritische Kernel-Updates, die einen Neustart erfordern, nicht automatisch herunter und installieren sie. Dies ist dasselbe Verhalten wie bei anderen Amazon-EC2-Instances, die das standardmäßige Amazon-Linux-AMI ausführen. Wenn neue Amazon-Linux-Softwareupdates, die einen Neustart erfordern (wie Kernel-, NVIDIA- und CUDA-Updates), nach der Veröffentlichung einer Amazon-EMR-Version verfügbar werden, laden EMR-Cluster-Instances, die das Standard-AMI ausführen, diese Updates nicht automatisch herunter und installieren sie. Um Kernel-Updates zu erhalten, können Sie [Ihr Amazon-EMR-AMI](#) so anpassen, dass es [das neueste Amazon-Linux-AMI verwendet](#).

- Die Konsolenunterstützung zur Erstellung einer Sicherheitskonfiguration, die die AWS Ranger-Integrationsoption spezifiziert, wird derzeit in der GovCloud Region nicht unterstützt. Die Sicherheitskonfiguration kann mit der CLI durchgeführt werden. Weitere Informationen finden [Sie unter Erstellen der EMR-Sicherheitskonfiguration](#) im Verwaltungshandbuch für Amazon EMR.
- Wenn AtRestEncryption oder HDFS-Verschlüsselung auf einem Cluster aktiviert ist, der Amazon EMR 5.31.0 oder 5.32.0 verwendet, führen Hive-Abfragen zu der folgenden Laufzeitausnahme.

```
TaskAttempt 3 failed, info=[Error: Error while running task ( failure ) :
attempt_1604112648850_0001_1_01_000000_3:java.lang.RuntimeException:
java.lang.RuntimeException: Hive Runtime Error while closing
operators: java.io.IOException: java.util.ServiceConfigurationError:
org.apache.hadoop.security.token.TokenIdentifier: Provider
org.apache.hadoop.hbase.security.token.AuthenticationTokenIdentifier not found
```

- Wenn Sie Spark mit der Formatierung des Hive-Partitionsstandorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den Amazon-EMR-Versionen 5.30.0 bis 5.36.0 und 6.2.0 bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:
  - Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.
  - Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist `s3://bucket/table/p=a` ein Präfix von `s3://bucket/table/p=a b`.

- Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen UTF-8-Wert, der kleiner als das /-Zeichen (U+002F) ist. Beispielsweise fällt das Leerzeichen (U+0020), das in `s3://bucket/table/p=a b` zwischen a und b vorkommt, in diese Kategorie. Beachten Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: `!"#$%&'()*+,-.` Weitere Informationen finden Sie unter [UTF-8-Kodierungstabelle und Unicode-Zeichen](#).

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.

## Version 6.2.0

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 6.2.0. Änderungen beziehen sich auf Version 6.1.0.

Veröffentlichungsdatum: 9. Dezember 2020

Letzte Aktualisierung: 4. Oktober 2021

### Unterstützte Anwendungen

- AWS SDK for Java Version 1.11.828
- emr-record-server Ausführung 1.7.0
- Flink-Version 1.11.2
- Ganglia Version 3.7.2
- Hadoop-Version 3.2.1-amzn-1
- HBase Version 2.2.6-amzn-0
- HB 1.0.0 ase-operator-tools
- HCatalog-Version 3.1.2-amzn-0
- Hive Version 3.1.2-amzn-3
- Hudi Version 0.6.0-amzn-1
- Hue Version 4.8.0
- JupyterHub Ausführung 1.1.0
- Livy-Version 0.7.0
- MXNet-Version 1.7.0

- Oozie-Version 5.2.0
- Phoenix Version 5.0.0
- Pig Version 0.17.0
- Presto Version 0.238.3-amzn-1
- PrestoSQL Version 343
- Spark Version 3.0.1-amzn-0
- Spark-Rapids 0.2.0
- TensorFlow Ausführung 2.3.1
- Zeppelin-Version 0.9.0-preview1
- Zookeeper Version 3.4.14
- Konnektoren und Treiber: DynamoDB Connector 4.16.0

## Neue Features

- HBase: Die Umbenennung in der Commit-Phase wurde entfernt und persistentes HFile-Tracking hinzugefügt. Weitere Informationen finden Sie unter [Persistentes HFile-Tracking](#) im Amazon-EMR-Versionshandbuch.
- HBase: Backported [Erstellen Sie eine Konfiguration, die das Zwischenspeichern von Blöcken bei der Komprimierung erzwingt.](#)
- PrestoDB: Verbesserungen beim dynamischen Partitionsbereinigen. Die regelbasierte Join Reorder funktioniert mit nicht partitionierten Daten.
- Verwaltete Richtlinien mit Geltungsbereich: Um den AWS bewährten Methoden zu entsprechen, hat Amazon EMR verwaltete Standardrichtlinien mit EMR-Geltungsbereich der Version v2 eingeführt, die als Ersatz für Richtlinien dienen, die nicht mehr unterstützt werden. Weitere Informationen finden Sie unter [Verwaltete Richtlinien von Amazon EMR](#).
- Unterstützungsstatus des Instance Metadata Service (IMDS) V2: Für Amazon EMR 6.2 oder höher verwenden Amazon-EMR-Komponenten IMDSv2 für alle IMDS-Aufrufe. Für IMDS-Aufrufe in Ihrem Anwendungscode können Sie sowohl IMDSv1 als auch IMDSv2 verwenden oder das IMDS so konfigurieren, dass es aus Sicherheitsgründen nur IMDSv2 verwendet. Wenn Sie IMDSv1 in früheren Amazon-EMR-6.x-Versionen deaktivieren, führt dies zu einem Cluster-Startup-Fehler.

## Änderungen, Verbesserungen und behobene Probleme

- Dies ist eine Version zur Behebung von Problemen mit Amazon-EMR-Skalierung, wenn ein Cluster nicht erfolgreich hoch-/herunterskaliert werden kann oder Anwendungsfehler verursacht werden.
- Es wurde ein Problem behoben, bei dem Skalierungsanforderungen für einen großen, stark ausgelasteten Cluster fehlschlagen, wenn Amazon-EMR-On-Cluster-Daemons Aktivitäten zur Integritätsprüfung durchführten, z. B. das Erfassen des YARN-Knotenstatus und des HDFS-Knotenstatus. Das lag daran, dass Cluster-Daemons die Integritätsstatusdaten eines Knotens nicht an interne Amazon-EMR-Komponenten weitergeben konnten.
- Verbesserte EMR-On-Cluster-Daemons zur korrekten Nachverfolgung der Knotenstatus bei der Wiederverwendung von IP-Adressen, um die Zuverlässigkeit bei Skalierungsvorgängen zu verbessern.
- [SPARK-29683](#). Es wurde ein Problem behoben, bei dem während der Cluster-Skalierung Auftragsfehler auftraten, da Spark davon ausging, dass alle verfügbaren Knoten auf der Verweigern-Liste standen.
- [YARN-9011](#). Es wurde ein Problem behoben, bei dem Auftragsfehler aufgrund eines Fehlers bei der Außerbetriebnahme von YARN auftraten, wenn der Cluster versuchte, hoch- oder herunterzuskalieren.
- Das Problem mit Schritt- oder Auftragsfehlern bei der Cluster-Skalierung wurde behoben, indem sichergestellt wurde, dass die Knotenstatus zwischen den Amazon-EMR-On-Cluster-Daemons und YARN/HDFS immer konsistent sind.
- Es wurde ein Problem behoben, bei dem Clustervorgänge wie Herunterskalierung und Schrittübermittlung für Amazon-EMR-Cluster, die mit Kerberos-Authentifizierung aktiviert waren, fehlschlagen. Dies lag daran, dass der Amazon-EMR-On-Cluster-Daemon das Kerberos-Ticket nicht erneuert hat, das für die sichere Kommunikation mit HDFS/YARN erforderlich ist, das auf dem Primärknoten ausgeführt wird.
- Neuere Amazon-EMR-Versionen beheben das Problem mit einem niedrigeren Limit für die „maximale Anzahl geöffneter Dateien“ auf älteren AL2-Versionen in Amazon EMR. Die Amazon-EMR-Versionen 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 und höher enthalten jetzt einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“.
- Spark: Leistungsverbesserungen in der Spark-Laufzeit.

## Bekannte Probleme

- Amazon EMR 6.2 hat falsche Berechtigungen für die Datei `libinstance-controller-java /etc/cron.d/` in EMR 6.2.0 festgelegt. Die Berechtigungen für die Datei sind `645 (-rw-r--r-x)`, obwohl sie `644 (-rw-r--r--)` sein sollten. Aus diesem Grund protokolliert Amazon-EMR-Version 6.2 keine Instance-State-Logs und das Verzeichnis `/emr/instance-logs` ist leer. Dieses Problem wurde in Amazon EMR 6.3.0 und höher behoben.

Zur Umgehung dieses Problems führen Sie das folgende Skript als Bootstrap-Aktion beim Clusterstart aus.

```
#!/bin/bash
sudo chmod 644 /etc/cron.d/libinstance-controller-java
```

- Für private Subnetz-Cluster von Amazon EMR 6.2.0 und 6.3.0 können Sie nicht auf die Ganglia-Webbenutzeroberfläche zugreifen. Sie erhalten die Fehlermeldung „Zugriff verweigert (403)“. Andere Web-Benutzeroberflächen wie Spark, Hue, Zeppelin JupyterHub, Livy und Tez funktionieren normal. Der Zugriff auf die Ganglia-Web-Benutzeroberfläche auf öffentlichen Subnetzclustern funktioniert ebenfalls normal. Um dieses Problem zu beheben, starten Sie den `httpd`-Service auf dem Primärknoten mit `sudo systemctl restart httpd` neu. Dieses Problem wurde in Amazon EMR 6.4.0 behoben.
- In Amazon EMR 6.2.0 gibt es ein Problem, bei dem `httpd` kontinuierlich ausfällt, wodurch Ganglia nicht verfügbar ist. Sie erhalten die Fehlermeldung „Es kann keine Verbindung zum Server hergestellt werden“. Um einen Cluster zu reparieren, der bereits mit diesem Problem läuft, stellen Sie eine SSH-Verbindung zum Cluster-Primärknoten her und fügen Sie die Zeile `Listen 80` zu der Datei `httpd.conf` hinzu, die sich unter `/etc/httpd/conf/httpd.conf` befindet. Dieses Problem wurde in Amazon EMR 6.3.0 behoben.
- `HTTPD` schlägt auf EMR-6.2.0-Clustern fehl, wenn Sie eine Sicherheitskonfiguration verwenden. Dadurch ist die Benutzeroberfläche der Ganglia-Webanwendung nicht verfügbar. Um auf die Benutzeroberfläche der Ganglia-Webanwendung zuzugreifen, fügen Sie `Listen 80` der `/etc/httpd/conf/httpd.conf`-Datei auf dem Primärknoten Ihres Clusters etwas hinzu. Informationen zum Herstellen einer Verbindung zu Ihrem Cluster finden Sie unter [Verbinden mit dem Primärknoten über SSH](#).

EMR Notebooks können auch keine Verbindung mit EMR-6.2.0-Clustern herstellen, wenn Sie eine Sicherheitskonfiguration verwenden. Das Notebook kann keine Kernel auflisten und Spark-



Aufträge nicht weiterleiten. Wir empfehlen, stattdessen EMR Notebooks mit einer anderen Version von Amazon EMR zu verwenden.

- Niedrigeres Limit für die „Max. Anzahl geöffneter Dateien“ auf älteren AL2-Versionen [in neueren Versionen behoben]. Amazon-EMR-Versionen: emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0 und emr-6.2.0 basieren auf älteren Versionen von Amazon Linux 2 (AL2), die eine niedrigere Ulimit-Einstellung für „Max. Anzahl geöffneter Dateien“ haben, wenn Amazon-EMR-Cluster mit dem Standard-AMI erstellt werden. Die Amazon-EMR-Versionen 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 und höher enthalten einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“. Versionen mit einem niedrigeren Limit für geöffnete Dateien verursachen beim Senden des Spark-Jobs den Fehler „Zu viele offene Dateien“. In den betroffenen Versionen hat das Amazon-EMR-Standard-AMI eine Ulimit-StandardEinstellung von 4 096 für „Max. Anzahl geöffneter Dateien“, was unter dem Dateilimit von 65 536 im neuesten Amazon-Linux-2-AMI liegt. Die niedrigere Ulimit-Einstellung für „Max. Anzahl geöffneter Dateien“ führt dazu, dass der Spark-Job fehlschlägt, wenn der Spark-Treiber und der Executor versuchen, mehr als 4 096 Dateien zu öffnen. Um das Problem zu beheben, verfügt Amazon EMR über ein Bootstrap Action (BA)-Skript, das die Ulimit-Einstellung bei der Cluster-Erstellung anpasst.

Wenn Sie eine ältere Amazon-EMR-Version verwenden, für die dieses Problem nicht dauerhaft behoben ist, können Sie mit der folgenden Problemumgehung das Instance-Controller-Ulimit explizit auf maximal 65 536 Dateien festlegen.

Explizit ein ulimit über die Befehlszeile setzen

1. Bearbeiten Sie `/etc/systemd/system/instance-controller.service`, um die folgenden Parameter zum Abschnitt Service hinzuzufügen.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Starten Sie neu InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Mithilfe der Bootstrap-Aktion (BA) ein Ulimit festlegen

Sie können auch ein Bootstrap-Aktionsskript (BA) verwenden, um das Ulimit für den Instance-Controller bei der Clustererstellung auf 65 536 Dateien zu konfigurieren.

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

• **⚠ Important**

Amazon EMR 6.1.0 und 6.2.0 beinhalten ein Leistungsproblem, das sich kritisch auf alle Hudi-Operationen zum Einfügen, Upsert und Löschen auswirken kann. Wenn Sie Hudi mit Amazon EMR 6.1.0 oder 6.2.0 verwenden möchten, sollten Sie sich an den AWS Support wenden, um ein gepatchtes Hudi RPM zu erhalten.

• **⚠ Important**

EMR-Cluster, auf denen Amazon-Linux- oder Amazon-Linux-2-AMIs (Amazon Machine Images) ausgeführt werden, verwenden das Standardverhalten von Amazon Linux und laden wichtige und kritische Kernel-Updates, die einen Neustart erfordern, nicht automatisch herunter und installieren sie. Dies ist dasselbe Verhalten wie bei anderen Amazon-EC2-Instances, die das standardmäßige Amazon-Linux-AMI ausführen. Wenn neue Amazon-Linux-Softwareupdates, die einen Neustart erfordern (wie Kernel-, NVIDIA- und CUDA-Updates), nach der Veröffentlichung einer Amazon-EMR-Version verfügbar werden, laden EMR-Cluster-Instances, die das Standard-AMI ausführen, diese Updates

nicht automatisch herunter und installieren sie. Um Kernel-Updates zu erhalten, können Sie [Ihr Amazon-EMR-AMI](#) so anpassen, dass es [das neueste Amazon-Linux-AMI verwendet](#).

- Maven-Artefakte von Amazon EMR 6.2.0 werden nicht veröffentlicht. Sie werden mit einer zukünftigen Version von Amazon EMR veröffentlicht.
- Die persistente HFile-Nachverfolgung mithilfe der HBase-Storefile-Systemtabelle unterstützt das Feature zur Replikation von HBase-Regionen nicht. Weitere Informationen zur HBase-Regionsreplikation finden Sie unter [Zeitleistenkonsistente, hochverfügbare Lesevorgänge](#).
- Unterschiede zwischen den Bucketing-Versionen von Amazon EMR 6.x und EMR 5.x Hive

EMR 5.x verwendet OOS Apache Hive 2, während in EMR 6.x OOS Apache Hive 3 verwendet wird. Die Open-Source-Version Hive2 verwendet Bucketing Version 1, während die Open-Source-Version Hive3 Bucketing Version 2 verwendet. Dieser Unterschied in der Bucketing-Version zwischen Hive 2 (EMR 5.x) und Hive 3 (EMR 6.x) bedeutet, dass Hive-Bucketing-Hashing anders funktioniert. Sehen Sie sich das folgende Beispiel an.

Die folgende Tabelle ist ein Beispiel, das in EMR 6.x bzw. EMR 5.x erstellt wurde.

```
-- Using following LOCATION in EMR 6.x
CREATE TABLE test_bucketing (id INT, desc STRING)
PARTITIONED BY (day STRING)
CLUSTERED BY(id) INTO 128 BUCKETS
LOCATION 's3://your-own-s3-bucket/emr-6-bucketing/';

-- Using following LOCATION in EMR 5.x
LOCATION 's3://your-own-s3-bucket/emr-5-bucketing/';
```

Dieselben Daten werden sowohl in EMR 6.x als auch in EMR 5.x eingefügt.

```
INSERT INTO test_bucketing PARTITION (day='01') VALUES(66, 'some_data');
INSERT INTO test_bucketing PARTITION (day='01') VALUES(200, 'some_data');
```

Die Überprüfung des S3-Speicherorts zeigt, dass der Name der Bucketing-Datei unterschiedlich ist, da sich die Hashing-Funktion zwischen EMR 6.x (Hive 3) und EMR 5.x (Hive 2) unterscheidet.

```
[hadoop@ip-10-0-0-122 ~]$ aws s3 ls s3://your-own-s3-bucket/emr-6-bucketing/day=01/
2020-10-21 20:35:16          13 000025_0
2020-10-21 20:35:22          14 000121_0
[hadoop@ip-10-0-0-122 ~]$ aws s3 ls s3://your-own-s3-bucket/emr-5-bucketing/day=01/
```

```
2020-10-21 20:32:07      13 000066_0
2020-10-21 20:32:51      14 000072_0
```

Sie können den Versionsunterschied auch erkennen, indem Sie den folgenden Befehl in der Hive-CLI in EMR 6.x ausführen. Beachten Sie, dass die Bucketing-Version 2 zurückgegeben wird.

```
hive> DESCRIBE FORMATTED test_bucketing;
...
Table Parameters:
  bucketing_version      2
...
```

- **Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung**

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in Amazon-EMR-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie dem Herunterskalieren oder der schrittweisen Übermittlung auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

Workaround:

- SSH als hadoop-Benutzer für den führenden Primärknoten des EMR-Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den hadoop-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

### Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen

konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

- Wenn Sie Spark mit der Formatierung des Hive-Partitionsstandorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den Amazon-EMR-Versionen 5.30.0 bis 5.36.0 und 6.2.0 bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:
  - Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.
  - Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist `s3://bucket/table/p=a` ein Präfix von `s3://bucket/table/p=a b`.
  - Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen UTF-8-Wert, der kleiner als das `/`-Zeichen (U+002F) ist. Beispielsweise fällt das Leerzeichen (U+0020), das in `s3://bucket/table/p=a b` zwischen `a` und `b` vorkommt, in diese Kategorie. Beachten Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: `!"#$%&'()*+,-.` Weitere Informationen finden Sie unter [UTF-8-Kodierungstabelle und Unicode-Zeichen](#).

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.

## Version 5.31.0

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.31.0. Änderungen beziehen sich auf Version 5.30.1.

Erste Version: 09. Oktober 2020

Letzte Aktualisierung: 15. Oktober 2020

### Upgrades

- Der Amazon-Glue-Konnektor wurde auf Version 1.13.0 aktualisiert
- Amazon SageMaker Spark SDK wurde auf Version 1.4.0 aktualisiert
- Der Amazon-Kinesis-Konnektor wurde auf Version 3.5.9 aktualisiert
- Auf Version AWS SDK for Java 1.11.852 aktualisiert
- BigTop-Tomcat wurde auf Version 8.5.56 aktualisiert

- EMR FS wurde auf Version 2.43.0 aktualisiert
- MetricsAndEventsApiGateway EMR-Client auf Version 1.4.0 aktualisiert
- EMR S3 Dist CP wurde auf Version 2.15.0 aktualisiert
- EMR S3 Select wurde auf Version 1.6.0 aktualisiert
- Flink wurde auf Version 1.11.0 aktualisiert
- Hadoop wurde auf Version 2.10.0 aktualisiert
- Hive wurde auf Version 2.3.7 aktualisiert
- Hudi wurde auf Version 0.6.0 aktualisiert
- Hue wurde auf Version 4.7.1 aktualisiert
- Auf Version JupyterHub 1.1.0 aktualisiert
- Mxnet wurde auf Version 1.6.0 aktualisiert
- OpenCV wurde auf Version 4.3.0 aktualisiert
- Presto wurde auf Version 0.238.3 aktualisiert
- Auf Version TensorFlow 2.1.0 aktualisiert

#### Änderungen, Verbesserungen und behobene Probleme

- Dies ist eine Version zur Behebung von Problemen mit Amazon-EMR-Skalierung, wenn ein Cluster nicht erfolgreich hoch-/herunterskaliert werden kann oder Anwendungsfehler verursacht werden.
- Es wurde ein Problem behoben, bei dem Skalierungsanforderungen für einen großen, stark ausgelasteten Cluster fehlschlugen, wenn Amazon-EMR-On-Cluster-Daemons Aktivitäten zur Integritätsprüfung durchführten, z. B. das Erfassen des YARN-Knotenstatus und des HDFS-Knotenstatus. Das lag daran, dass Cluster-Daemons die Integritätsstatusdaten eines Knotens nicht an interne Amazon-EMR-Komponenten weitergeben konnten.
- Verbesserte EMR-On-Cluster-Daemons zur korrekten Nachverfolgung der Knotenstatus bei der Wiederverwendung von IP-Adressen, um die Zuverlässigkeit bei Skalierungsvorgängen zu verbessern.
- [SPARK-29683](#). Es wurde ein Problem behoben, bei dem während der Cluster-Skalierung Auftragsfehler auftraten, da Spark davon ausging, dass alle verfügbaren Knoten auf der Verweigern-Liste standen.
- [YARN-9011](#). Es wurde ein Problem behoben, bei dem Auftragsfehler aufgrund eines Fehlers bei der Außerbetriebnahme von YARN auftraten, wenn der Cluster versuchte, hoch- oder herunterzuskalieren.

- Das Problem mit Schritt- oder Auftragsfehlern bei der Cluster-Skalierung wurde behoben, indem sichergestellt wurde, dass die Knotenstatus zwischen den Amazon-EMR-On-Cluster-Daemons und YARN/HDFS immer konsistent sind.
- Es wurde ein Problem behoben, bei dem Clustervorgänge wie Herunterskalierung und Schrittübermittlung für Amazon-EMR-Cluster, die mit Kerberos-Authentifizierung aktiviert waren, fehlschlagen. Dies lag daran, dass der Amazon-EMR-On-Cluster-Daemon das Kerberos-Ticket nicht erneuert hat, das für die sichere Kommunikation mit HDFS/YARN erforderlich ist, das auf dem Primärknoten ausgeführt wird.
- Neuere Amazon-EMR-Versionen beheben das Problem mit einem niedrigeren Limit für die „maximale Anzahl geöffneter Dateien“ auf älteren AL2-Versionen in Amazon EMR. Die Amazon-EMR-Versionen 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 und höher enthalten jetzt einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“.
- [Hive Spaltenstatistiken](#) werden für Amazon-EMR-5.31.0 und höher unterstützt.
- Aktualisierte Komponentenversionen.
- Support für EMRFS S3EC V2 in Amazon EMR 5.31.0. In den S3-Java-SDK-Versionen 1.11.837 und höher wurde der Verschlüsselungsclient Version 2 (S3EC V2) mit verschiedenen Sicherheitsverbesserungen eingeführt. Weitere Informationen finden Sie hier:
  - S3 Blogbeitrag: [Updates für den Amazon S3 Encryption Client](#).
  - AWS SDK for Java Entwicklerhandbuch: [Migrieren Sie Verschlüsselungs- und Entschlüsselungsclients auf V2](#).
  - EMR-Verwaltungshandbuch: [Amazon-S3-Clientseitige Verschlüsselung](#).

Der Encryption Client V1 ist aus Gründen der Abwärtskompatibilität weiterhin im SDK verfügbar.

## Neue Features

- Niedrigeres Limit für die „Max. Anzahl geöffneter Dateien“ auf älteren AL2-Versionen [in neueren Versionen behoben]. Amazon-EMR-Versionen: emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0 und emr-6.2.0 basieren auf älteren Versionen von Amazon Linux 2 (AL2), die eine niedrigere Ulimit-Einstellung für „Max. Anzahl geöffneter Dateien“ haben, wenn Amazon-EMR-Cluster mit dem Standard-AMI erstellt werden. Die Amazon-EMR-Versionen 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 und höher enthalten einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“. Versionen mit einem niedrigeren Limit für geöffnete Dateien verursachen beim Senden des Spark-Jobs den Fehler „Zu viele offene Dateien“. In den betroffenen Versionen hat das Amazon-EMR-Standard-AMI eine Ulimit-StandardEinstellung

von 4 096 für „Max. Anzahl geöffneter Dateien“, was unter dem Dateilimit von 65 536 im neuesten Amazon-Linux-2-AMI liegt. Die niedrigere Ulimit-Einstellung für „Max. Anzahl geöffneter Dateien“ führt dazu, dass der Spark-Job fehlschlägt, wenn der Spark-Treiber und der Executor versuchen, mehr als 4 096 Dateien zu öffnen. Um das Problem zu beheben, verfügt Amazon EMR über ein Bootstrap Action (BA)-Skript, das die Ulimit-Einstellung bei der Cluster-Erstellung anpasst.

Wenn Sie eine ältere Amazon-EMR-Version verwenden, für die dieses Problem nicht dauerhaft behoben ist, können Sie mit der folgenden Problemumgehung das Instance-Controller-Ulimit explizit auf maximal 65 536 Dateien festlegen.

Explizit ein ulimit über die Befehlszeile setzen

1. Bearbeiten Sie `/etc/systemd/system/instance-controller.service`, um die folgenden Parameter zum Abschnitt Service hinzuzufügen.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Starten Sie neu InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Mithilfe der Bootstrap-Aktion (BA) ein Ulimit festlegen

Sie können auch ein Bootstrap-Aktionsskript (BA) verwenden, um das Ulimit für den Instance-Controller bei der Clustererstellung auf 65 536 Dateien zu konfigurieren.

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
```



```
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

- Mit Amazon EMR 5.31.0 können Sie einen Cluster starten, der in Lake Formation integriert ist. Diese Integration ermöglicht eine detaillierte Datenfilterung auf Spaltenebene für Datenbanken und Tabellen im Glue-Datenkatalog. AWS ermöglicht Lake Formation über ein Unternehmens-Identitätssystem eine verbundene einmalige Anmeldung bei EMR Notebooks oder Apache Zeppelin. Weitere Informationen über [Integrieren von Amazon EMR mit AWS Lake Formation](#) finden Sie im Verwaltungshandbuch für Amazon EMR.

Amazon EMR with Lake Formation ist derzeit in 16 AWS Regionen verfügbar: USA Ost (Ohio und Nord-Virginia), USA West (Nordkalifornien und Oregon), Asien-Pazifik (Mumbai, Seoul, Singapur, Sydney und Tokio), Kanada (Zentral), Europa (Frankfurt, Irland, London, Paris und Stockholm), Südamerika (São Paulo).

## Bekannte Probleme

- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in Amazon-EMR-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie dem Herunterskalieren oder der schrittweisen Übermittlung auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

### Workaround:

- SSH als hadoop-Benutzer für den führenden Primärknoten des EMR-Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den hadoop-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

### Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

- Wenn AtRestEncryption oder HDFS-Verschlüsselung auf einem Cluster aktiviert ist, der Amazon EMR 5.31.0 oder 5.32.0 verwendet, führen Hive-Abfragen zu der folgenden Laufzeitausnahme.

```
TaskAttempt 3 failed, info=[Error: Error while running task ( failure ) :
attempt_1604112648850_0001_1_01_000000_3:java.lang.RuntimeException:
java.lang.RuntimeException: Hive Runtime Error while closing
operators: java.io.IOException: java.util.ServiceConfigurationError:
org.apache.hadoop.security.token.TokenIdentifier: Provider
org.apache.hadoop.hbase.security.token.AuthenticationTokenIdentifier not found
```

- Wenn Sie Spark mit der Formatierung des Hive-Partitionsstandorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den Amazon-EMR-Versionen 5.30.0 bis 5.36.0 und 6.2.0 bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:
  - Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.
  - Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist `s3://bucket/table/p=a` ein Präfix von `s3://bucket/table/p=a b`.
  - Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen UTF-8-Wert, der kleiner als das `/`-Zeichen (U+002F) ist. Beispielsweise fällt das Leerzeichen (U+0020), das in `s3://bucket/table/p=a b` zwischen `a` und `b` vorkommt, in diese Kategorie. Beachten Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: `!"#$%&'()*+,-.` Weitere Informationen finden Sie unter [UTF-8-Kodierungstabelle und Unicode-Zeichen](#).

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.

## Version 6.1.0

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 6.1.0. Änderungen beziehen sich auf Version 6.0.0.

Erste Version: 4. September 2020

Letzte Aktualisierung: 15. Oktober 2020

### Unterstützte Anwendungen

- AWS SDK for Java Version 1.11.828
- Flink-Version 1.11.0
- Ganglia Version 3.7.2
- Hadoop-Version 3.2.1-amzn-1
- HBase-Version 2.2.5
- ase-operator-tools HB 1.0.0
- HCatalog-Version 3.1.2-amzn-0
- Hive-Version 3.1.2-amzn-1
- Hudi-Version 0.5.2-incubating
- Hue-Version 4.7.1
- JupyterHub Ausführung 1.1.0
- Livy-Version 0.7.0
- MXNet-Version 1.6.0
- Oozie-Version 5.2.0
- Phoenix Version 5.0.0
- Presto-Version 0.232
- PrestoSQL-Version 338
- Spark-Version 3.0.0-amzn-0
- TensorFlow Ausführung 2.1.0
- Zeppelin-Version 0.9.0-preview1
- Zookeeper Version 3.4.14
- Anschlüsse und Treiber: DynamoDB Connector 4.14.0

## Neue Features

- ARM-Instance-Typen werden ab Amazon-EMR-Version 5.30.0 und Amazon-EMR-Version 6.1.0 unterstützt.
- Allzweck-Instance-Typen von M6g werden ab den Amazon-EMR-Versionen 6.1.0 und 5.30.0 unterstützt. Weitere Informationen finden Sie unter [Instance-Flotten konfigurieren](#) im Verwaltungshandbuch für Amazon EMR.
- Das EC2-Platzierungsgruppenfeature wird ab Amazon-EMR-Version 5.23.0 als Option für Cluster mit mehreren Primärknoten unterstützt. Derzeit werden nur Primärknoten-Typen von des Platzierungsgruppenfeatures unterstützt, und die SPREAD-Strategie wird auf diese Primärknoten angewendet. Bei dieser SPREAD-Strategie wird eine kleine Gruppe von Instances auf separater zugrundeliegender Hardware platziert, um den Verlust mehrerer Primärknoten im Falle eines Hardwarefehlers zu verhindern. Weitere Informationen finden Sie unter [EMR-Integration mit EC2 Placement Group](#) im Verwaltungshandbuch für Amazon EMR.
- **Verwaltete Skalierung** – Mit Amazon EMR ab Version 6.1.0 können Sie Amazon EMR Managed Scaling aktivieren, um die Anzahl der Instances oder Einheiten in Ihrem Cluster basierend auf der Workload automatisch zu erhöhen oder zu verringern. Amazon EMR wertet Cluster-Metriken kontinuierlich aus, um Skalierungsentscheidungen zu treffen, die Ihre Cluster für Kosten und Geschwindigkeit optimieren. Verwaltete Skalierung ist auch in Amazon-EMR-Version 5.30.0 und höher verfügbar, außer 6.0.0. Weitere Informationen finden Sie unter [Skalieren von Clusterressourcen](#) im Verwaltungshandbuch für Amazon EMR.
- **PrestoSQL-Version 338** wird mit EMR 6.1.0 unterstützt. Weitere Informationen finden Sie unter [Presto](#).
  - PrestoSQL wird nur auf EMR 6.1.0 und späteren Versionen unterstützt, nicht auf EMR 6.0.0 oder EMR 5.x.
  - Der Anwendungsname `Presto` wird weiterhin verwendet, um PrestoDB auf Clustern zu installieren. Verwenden Sie den Anwendungsnamen `PrestoSQL`, um PrestoSQL auf Clustern zu installieren.
  - Sie können entweder PrestoDB oder PrestoSQL installieren, aber Sie können nicht beide auf einem einzigen Cluster installieren. Wenn beim Versuch, einen Cluster zu erstellen, sowohl PrestoDB als auch PrestoSQL angegeben werden, tritt ein Validierungsfehler auf und die Clustererstellungsanforderung schlägt fehl.
  - PrestoSQL wird sowohl auf Single-Haupt- als auch auf Multi-Haupt-Clustern unterstützt. Auf Multi-Haupt-Clustern ist ein externer Hive-Metastore erforderlich, um PrestoSQL oder PrestoDB

auszuführen. Weitere Informationen finden Sie unter [Unterstützte Anwendungen in einem EMR-Cluster mit mehreren Primärknoten](#).

- Unterstützung der auto ECR-Authentifizierung auf Apache Hadoop und Apache Spark mit Docker: Spark-Benutzer können Docker-Images von Docker Hub und Amazon Elastic Container Registry (Amazon ECR) verwenden, um Umgebungs- und Bibliotheksabhängigkeiten zu definieren.

[Konfigurieren Sie Docker](#) und [führen Sie Spark-Anwendungen mit Docker mithilfe von Amazon EMR 6.x aus](#).

- EMR unterstützt Apache Hive ACID-Transaktionen: Amazon EMR 6.1.0 bietet Unterstützung für Hive-ACID-Transaktionen, sodass es den ACID-Eigenschaften einer Datenbank entspricht. Mit diesem Feature können Sie INSERT, UPDATE, DELETE, - und MERGE-Operationen in Hive-Tabellen mit Daten in Amazon Simple Storage Service (Amazon S3) ausführen. Dies ist ein wichtiges Feature für Anwendungsfälle wie Streaming-Erfassung, Neuformulierung von Daten, Massenaktualisierungen mit MERGE und sich langsam ändernde Dimensionen. Weitere Informationen, einschließlich Konfigurationsbeispielen und Anwendungsfällen, finden Sie unter [Amazon EMR unterstützt Apache-Hive-ACID-Transaktionen](#).

## Änderungen, Verbesserungen und behobene Probleme

- Dies ist eine Version zur Behebung von Problemen mit Amazon-EMR-Skalierung, wenn ein Cluster nicht erfolgreich hoch-/herunterskaliert werden kann oder Anwendungsfehler verursacht werden.
- Es wurde ein Problem behoben, bei dem Skalierungsanforderungen für einen großen, stark ausgelasteten Cluster fehlschlagen, wenn Amazon-EMR-On-Cluster-Daemons Aktivitäten zur Integritätsprüfung durchführten, z. B. das Erfassen des YARN-Knotenstatus und des HDFS-Knotenstatus. Das lag daran, dass Cluster-Daemons die Integritätsstatusdaten eines Knotens nicht an interne Amazon-EMR-Komponenten weitergeben konnten.
- Verbesserte EMR-On-Cluster-Daemons zur korrekten Nachverfolgung der Knotenstatus bei der Wiederverwendung von IP-Adressen, um die Zuverlässigkeit bei Skalierungsvorgängen zu verbessern.
- [SPARK-29683](#). Es wurde ein Problem behoben, bei dem während der Cluster-Skalierung Auftragsfehler auftraten, da Spark davon ausging, dass alle verfügbaren Knoten auf der Verweigern-Liste standen.
- [YARN-9011](#). Es wurde ein Problem behoben, bei dem Auftragsfehler aufgrund eines Fehlers bei der Außerbetriebnahme von YARN auftraten, wenn der Cluster versuchte, hoch- oder herunterzuskalieren.

- Das Problem mit Schritt- oder Auftragsfehlern bei der Cluster-Skalierung wurde behoben, indem sichergestellt wurde, dass die Knotenstatus zwischen den Amazon-EMR-On-Cluster-Daemons und YARN/HDFS immer konsistent sind.
- Es wurde ein Problem behoben, bei dem Clustervorgänge wie Herunterskalierung und Schrittübermittlung für Amazon-EMR-Cluster, die mit Kerberos-Authentifizierung aktiviert waren, fehlschlagen. Dies lag daran, dass der Amazon-EMR-On-Cluster-Daemon das Kerberos-Ticket nicht erneuert hat, das für die sichere Kommunikation mit HDFS/YARN erforderlich ist, das auf dem Primärknoten ausgeführt wird.
- Neuere Amazon-EMR-Versionen beheben das Problem mit einem niedrigeren Limit für die „maximale Anzahl geöffneter Dateien“ auf älteren AL2-Versionen in Amazon EMR. Die Amazon-EMR-Versionen 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 und höher enthalten jetzt einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“.
- Apache Flink wird auf EMR 6.0.0 nicht unterstützt, aber es wird auf EMR 6.1.0 mit Flink 1.11.0 unterstützt. Dies ist die erste Version von Flink, die Hadoop 3 offiziell unterstützt. Siehe Ankündigung der Veröffentlichung von [Apache Flink 1.11.0](#).
- Ganglia wurde aus den standardmäßigen EMR-6.1.0-Paketen entfernt.

## Bekannte Probleme

- Niedrigeres Limit für die „Max. Anzahl geöffneter Dateien“ auf älteren AL2-Versionen [in neueren Versionen behoben]. Amazon-EMR-Versionen: emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0 und emr-6.2.0 basieren auf älteren Versionen von Amazon Linux 2 (AL2), die eine niedrigere Ulimit-Einstellung für „Max. Anzahl geöffneter Dateien“ haben, wenn Amazon-EMR-Cluster mit dem Standard-AMI erstellt werden. Die Amazon-EMR-Versionen 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 und höher enthalten einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“. Versionen mit einem niedrigeren Limit für geöffnete Dateien verursachen beim Senden des Spark-Jobs den Fehler „Zu viele offene Dateien“. In den betroffenen Versionen hat das Amazon-EMR-Standard-AMI eine Ulimit-StandardEinstellung von 4 096 für „Max. Anzahl geöffneter Dateien“, was unter dem Dateilimit von 65 536 im neuesten Amazon-Linux-2-AMI liegt. Die niedrigere Ulimit-Einstellung für „Max. Anzahl geöffneter Dateien“ führt dazu, dass der Spark-Job fehlschlägt, wenn der Spark-Treiber und der Executor versuchen, mehr als 4 096 Dateien zu öffnen. Um das Problem zu beheben, verfügt Amazon EMR über ein Bootstrap Action (BA)-Skript, das die Ulimit-Einstellung bei der Cluster-Erstellung anpasst.

Wenn Sie eine ältere Amazon-EMR-Version verwenden, für die dieses Problem nicht dauerhaft behoben ist, können Sie mit der folgenden Problemumgehung das Instance-Controller-Ulimit explizit auf maximal 65 536 Dateien festlegen.

Explizit ein ulimit über die Befehlszeile setzen

1. Bearbeiten Sie `/etc/systemd/system/instance-controller.service`, um die folgenden Parameter zum Abschnitt Service hinzuzufügen.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Starten Sie neu InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Mithilfe der Bootstrap-Aktion (BA) ein Ulimit festlegen

Sie können auch ein Bootstrap-Aktionsskript (BA) verwenden, um das Ulimit für den Instance-Controller bei der Clustererstellung auf 65 536 Dateien zu konfigurieren.

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
```

```
sudo systemctl daemon-reload
```

**⚠ Important**

Amazon EMR 6.1.0 und 6.2.0 beinhalten ein Leistungsproblem, das sich kritisch auf alle Hudi-Operationen zum Einfügen, Upsert und Löschen auswirken kann. Wenn Sie Hudi mit Amazon EMR 6.1.0 oder 6.2.0 verwenden möchten, sollten Sie sich an den AWS Support wenden, um ein gepatchtes Hudi RPM zu erhalten.

- Wenn Sie eine benutzerdefinierte Garbage-Collection-Konfiguration mit `spark.driver.extraJavaOptions` und `spark.executor.extraJavaOptions` festlegen, führt dies dazu, dass der Treiber/Executor-Start mit EMR 6.1 aufgrund einer widersprüchlichen Garbage-Collection-Konfiguration fehlschlägt. Mit EMR Version 6.1.0 sollten Sie stattdessen eine benutzerdefinierte Spark-Garbage-Collection-Konfiguration für Treiber und Executors mit den Eigenschaften `spark.driver.defaultJavaOptions` und `spark.executor.defaultJavaOptions` angeben. Weitere Informationen finden Sie unter [Apache Spark Runtime Environment](#) und [Konfigurieren von Spark Garbage Collection auf Amazon EMR 6.1.0](#).
- Die Verwendung von Pig mit Oozie (und innerhalb von Hue, da Hue Oozie-Aktionen zur Ausführung von Pig-Skripten verwendet) generiert den Fehler, dass eine native LZO-Bibliothek nicht geladen werden kann. Diese Fehlermeldung ist informativ und verhindert nicht, dass Pig ausgeführt wird.
- Hudi-Parallelitätsunterstützung: Derzeit unterstützt Hudi keine gleichzeitigen Schreibvorgänge in eine einzelne Hudi-Tabelle. Darüber hinaus macht Hudi alle Änderungen rückgängig, die von in Bearbeitung befindlichen Autoren vorgenommen wurden, bevor ein neuer Writer beginnen kann. Gleichzeitige Schreibvorgänge können diesen Mechanismus stören und zu Wettlaufbedingungen führen, was zu Datenbeschädigungen führen kann. Sie sollten sicherstellen, dass im Rahmen Ihres Datenverarbeitungs-Workflows immer nur ein einziger Hudi-Writer mit einer Hudi-Tabelle arbeitet. Hudi unterstützt mehrere gleichzeitige Lesegeräte, die mit derselben Hudi-Tabelle arbeiten.
- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in Amazon-EMR-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie dem Herunterskalieren oder der schrittweisen Übermittlung auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische



als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

Workaround:

- SSH als hadoop-Benutzer für den führenden Primärknoten des EMR-Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den hadoop-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

#### Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

- In Amazon EMR 6.1.0 gibt es ein Problem, das Cluster betrifft, auf denen Presto ausgeführt wird. Nach einem längeren Zeitraum (Tage) kann der Cluster Fehler wie „su: failed to execute /bin/bash: Resource temporarily unavailable“ oder „Shell-Anfrage auf Kanal 0 fehlgeschlagen“ ausgeben. Dieses Problem wird durch einen internen Amazon EMR-Prozess (InstanceController) verursacht, der zu viele Light Weight Processes (LWP) erzeugt, was letztendlich dazu führt, dass der Hadoop-Benutzer sein Nproc-Limit überschreitet. Dadurch wird verhindert, dass der Benutzer zusätzliche Prozesse öffnet. Die Lösung für dieses Problem ist ein Upgrade auf EMR 6.2.0.

## Version 6.0.0


Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 6.0.0.

Erste Version: 10. März 2020

Unterstützte Anwendungen

- AWS SDK for Java Version 1.11.711

- Ganglia Version 3.7.2
- Hadoop Version 3.2.1
- HBase Version 2.2.3
- HCatalog Version 3.1.2
- Hive Version 3.1.2
- Hudi Version 0.5.0-incubating
- Hue Version 4.4.0
- JupyterHub Ausführung 1.0.0
- Livy Version 0.6.0
- MXNet-Version 1.5.1
- Oozie Version 5.1.0
- Phoenix Version 5.0.0
- Presto Version 0.230
- Spark Version 2.4.4
- TensorFlow Ausführung 1.14.0
- Zeppelin Version 0.9.0-SNAPSHOT
- Zookeeper Version 3.4.14
- Anschlüsse und Treiber: DynamoDB Connector 4.14.0

 Note

Flink, Sqoop, Pig und Mahout sind in Amazon-EMR-Version 6.0.0 nicht verfügbar.

## Neue Features

- YARN Docker Runtime Support - YARN-Anwendungen, wie etwa Spark-Aufgaben, können jetzt im Kontext eines Docker-Containers ausgeführt werden. Dadurch können Sie ganz einfach Abhängigkeiten in einem Docker-Image definieren, ohne dass benutzerdefinierte Bibliotheken auf dem Amazon-EMR-Cluster installiert werden müssen. Weitere Informationen finden Sie unter [Konfigurieren der Docker-Integration](#) und [Ausführen von Spark-Anwendungen mit Docker mit Amazon EMR 6.0.0](#).

- Unterstützung für Hive LLAP - Hive unterstützt jetzt den LLAP-Ausführungsmodus für eine verbesserte Abfrageleistung. Weitere Informationen finden Sie unter [Verwenden von Hive LLAP](#).

## Änderungen, Verbesserungen und behobene Probleme

- Dies ist eine Version zur Behebung von Problemen mit Amazon-EMR-Skalierung, wenn ein Cluster nicht erfolgreich hoch-/herunterskaliert werden kann oder Anwendungsfehler verursacht werden.
- Es wurde ein Problem behoben, bei dem Skalierungsanforderungen für einen großen, stark ausgelasteten Cluster fehlschlagen, wenn Amazon-EMR-On-Cluster-Daemons Aktivitäten zur Integritätsprüfung durchführten, z. B. das Erfassen des YARN-Knotenstatus und des HDFS-Knotenstatus. Das lag daran, dass Cluster-Daemons die Integritätsstatusdaten eines Knotens nicht an interne Amazon-EMR-Komponenten weitergeben konnten.
- Verbesserte EMR-On-Cluster-Daemons zur korrekten Nachverfolgung der Knotenstatus bei der Wiederverwendung von IP-Adressen, um die Zuverlässigkeit bei Skalierungsvorgängen zu verbessern.
- [SPARK-29683](#). Es wurde ein Problem behoben, bei dem während der Cluster-Skalierung Auftragsfehler auftraten, da Spark davon ausging, dass alle verfügbaren Knoten auf der Verweigern-Liste standen.
- [YARN-9011](#). Es wurde ein Problem behoben, bei dem Auftragsfehler aufgrund eines Fehlers bei der Außerbetriebnahme von YARN auftraten, wenn der Cluster versuchte, hoch- oder herunterzuskalieren.
- Das Problem mit Schritt- oder Auftragsfehlern bei der Cluster-Skalierung wurde behoben, indem sichergestellt wurde, dass die Knotenstatus zwischen den Amazon-EMR-On-Cluster-Daemons und YARN/HDFS immer konsistent sind.
- Es wurde ein Problem behoben, bei dem Clustervorgänge wie Herunterskalierung und Schrittübermittlung für Amazon-EMR-Cluster, die mit Kerberos-Authentifizierung aktiviert waren, fehlschlagen. Dies lag daran, dass der Amazon-EMR-On-Cluster-Daemon das Kerberos-Ticket nicht erneuert hat, das für die sichere Kommunikation mit HDFS/YARN erforderlich ist, das auf dem Primärknoten ausgeführt wird.
- Neuere Amazon-EMR-Versionen beheben das Problem mit einem niedrigeren Limit für die „maximale Anzahl geöffneter Dateien“ auf älteren AL2-Versionen in Amazon EMR. Die Amazon-EMR-Versionen 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 und höher enthalten jetzt einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“.
- Amazon Linux
  - Amazon Linux 2 ist das Betriebssystem für die EMR 6.x-Release-Serie.

- `systemd` wird für die Serviceverwaltung statt `upstart` in Amazon Linux 1 verwendet.
- Java Development Kit (JDK)
  - Coretto JDK 8 ist das Standard-JDK für die EMR-6.x-Release-Serie.
- Scala
  - Scala 2.12 wird mit Apache Spark und Apache Livy verwendet.
- Python 3
  - Python 3 ist jetzt die Standardversion von Python in EMR.
- YARN-Knotenbeschriftungen
  - Beginnend mit der Amazon-EMR-6.x-Release-Reihe ist das Feature YARN-Knotenbeschriftungen standardmäßig deaktiviert. Die Anwendungs-Master-Prozesse können standardmäßig sowohl auf Kern- als auch auf Aufgabenknoten ausgeführt werden. Sie können die Funktion YARN-Knotenbeschriftungen aktivieren, indem Sie folgende Eigenschaften konfigurieren: `yarn.node-labels.enabled` und `yarn.node-labels.am.default-node-label-expression`. Weitere Informationen finden Sie unter [Grundlegendes zu Primär-, Kern- und Aufgabenknoten](#).

## Bekannte Probleme

- Niedrigeres Limit für die „Max. Anzahl geöffneter Dateien“ auf älteren AL2-Versionen [in neueren Versionen behoben]. Amazon-EMR-Versionen: `emr-5.30.x`, `emr-5.31.0`, `emr-5.32.0`, `emr-6.0.0`, `emr-6.1.0` und `emr-6.2.0` basieren auf älteren Versionen von Amazon Linux 2 (AL2), die eine niedrigere `ulimit`-Einstellung für „Max. Anzahl geöffneter Dateien“ haben, wenn Amazon-EMR-Cluster mit dem Standard-AMI erstellt werden. Die Amazon-EMR-Versionen `5.30.1`, `5.30.2`, `5.31.1`, `5.32.1`, `6.0.1`, `6.1.1`, `6.2.1`, `5.33.0`, `6.3.0` und höher enthalten einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“. Versionen mit einem niedrigeren Limit für geöffnete Dateien verursachen beim Senden des Spark-Jobs den Fehler „Zu viele offene Dateien“. In den betroffenen Versionen hat das Amazon-EMR-Standard-AMI eine `ulimit`-Standardeinstellung von 4 096 für „Max. Anzahl geöffneter Dateien“, was unter dem Dateilimit von 65 536 im neuesten Amazon-Linux-2-AMI liegt. Die niedrigere `ulimit`-Einstellung für „Max. Anzahl geöffneter Dateien“ führt dazu, dass der Spark-Job fehlschlägt, wenn der Spark-Treiber und der Executor versuchen, mehr als 4 096 Dateien zu öffnen. Um das Problem zu beheben, verfügt Amazon EMR über ein Bootstrap Action (BA)-Skript, das die `ulimit`-Einstellung bei der Cluster-Erstellung anpasst.

Wenn Sie eine ältere Amazon-EMR-Version verwenden, für die dieses Problem nicht dauerhaft behoben ist, können Sie mit der folgenden Problemumgehung das Instance-Controller-Ulimit explizit auf maximal 65 536 Dateien festlegen.

Explizit ein ulimit über die Befehlszeile setzen

1. Bearbeiten Sie `/etc/systemd/system/instance-controller.service`, um die folgenden Parameter zum Abschnitt Service hinzuzufügen.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Starten Sie neu InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Mithilfe der Bootstrap-Aktion (BA) ein Ulimit festlegen

Sie können auch ein Bootstrap-Aktionsskript (BA) verwenden, um das Ulimit für den Instance-Controller bei der Clustererstellung auf 65 536 Dateien zu konfigurieren.

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
```

```
sudo systemctl daemon-reload
```

- Die interaktive Spark-Shell PySpark, einschließlich SparkR und Spark-Shell, unterstützt die Verwendung von Docker mit zusätzlichen Bibliotheken nicht.
- Um Python 3 mit Amazon-EMR-Version 6.0.0 zu verwenden, müssen Sie PATH zu `yarn.nodemanager.env-whitelist` hinzufügen.
- Die Funktion Live Long and Process (LLAP) wird nicht unterstützt, wenn Sie den AWS Glue-Datenkatalog als Metastore für Hive verwenden.
- Wenn Sie Amazon EMR 6.0.0 mit Spark- und Docker-Integration verwenden, müssen Sie die Instances in Ihrem Cluster mit demselben Instance-Typ und derselben Anzahl an EBS-Volumes konfigurieren, um Fehler beim Senden eines Spark-Auftrags mit Docker-Laufzeit zu vermeiden.
- [In Amazon EMR 6.0.0 ist der Speichermodus HBase auf Amazon S3 vom HBASE-24286.-Problem betroffen.](#) Der HBase-Master kann nicht initialisiert werden, wenn der Cluster mit vorhandenen S3-Daten erstellt wird.
- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in Amazon-EMR-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie dem Herunterskalieren oder der schrittweisen Übermittlung auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

Workaround:

- SSH als `hadoop`-Benutzer für den führenden Primärknoten des EMR-Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den `hadoop`-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

**Note**

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

## Version 5.30.1

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.30.1. Änderungen beziehen sich auf Version 5.30.0.

Erstveröffentlichungsdatum: 30. Juni 2020

Letzte Aktualisierung: 24. August 2020

### Änderungen, Verbesserungen und behobene Probleme

- Neuere Amazon-EMR-Versionen beheben das Problem mit einem niedrigeren Limit für die „maximale Anzahl geöffneter Dateien“ auf älteren AL2-Versionen in Amazon EMR. Die Amazon-EMR-Versionen 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 und höher enthalten jetzt einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“.
- Es wurde ein Problem behoben, bei dem der Instance-Controller-Prozess unendlich viele Prozesse ausgelöst hat.
- Es wurde ein Problem behoben, bei dem Hue keine Hive-Abfrage ausführen konnte, die Meldung „Datenbank ist gesperrt“ angezeigt wurde und die Ausführung von Abfragen verhindert wurde.
- Es wurde ein Spark-Problem behoben, durch das mehr Aufgaben gleichzeitig auf dem EMR-Cluster ausgeführt werden konnten.
- Es wurde ein Problem mit dem Jupyter Notebook behoben, das den Fehler „Zu viele Dateien öffnen“ auf dem Jupyter-Server verursachte.
- Ein Problem mit den Startzeiten von Clustern wurde behoben.

### Neue Features

- Persistente Anwendungsschnittstellen für die Tez-Benutzeroberfläche und den YARN-Timeline-Server sind mit den Amazon-EMR-Versionen 6.x und EMR-Version 5.30.1 und höher verfügbar.

Durch den Linkzugriff mit einem Klick auf den persistenten Anwendungsverlauf können Sie schnell auf den Jobverlauf zugreifen, ohne einen Webproxy über eine SSH-Verbindung einrichten zu müssen. Protokolle für aktive und beendete Cluster sind 30 Tage nach Ende der Anwendung verfügbar. Weitere Informationen finden Sie unter [Ansicht der Benutzeroberflächen für persistente Anwendungen](#) im Verwaltungshandbuch für Amazon EMR an.

- EMR-Notebook-Ausführungs-APIs sind verfügbar, um EMR-Notebooks über ein Skript oder eine Befehlszeile auszuführen. Durch die Möglichkeit, EMR-Notebook-Ausführungen ohne die AWS Konsole zu starten, zu stoppen, aufzulisten und zu beschreiben, können Sie ein EMR-Notebook programmgesteuert steuern. Mithilfe einer parametrisierten Notebook-Zelle können Sie verschiedene Parameterwerte an ein Notebook übergeben, ohne für jeden neuen Satz von Parameterwerten eine Kopie des Notebooks erstellen zu müssen. Siehe [EMR-API-Aktionen](#). Beispielcode finden Sie unter [Beispielbefehle zur programmatischen Ausführung von EMR Notebooks](#).

## Bekannte Probleme

- Niedrigeres Limit für die „Max. Anzahl geöffneter Dateien“ auf älteren AL2-Versionen [in neueren Versionen behoben]. Amazon-EMR-Versionen: emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0 und emr-6.2.0 basieren auf älteren Versionen von Amazon Linux 2 (AL2), die eine niedrigere Ulimit-Einstellung für „Max. Anzahl geöffneter Dateien“ haben, wenn Amazon-EMR-Cluster mit dem Standard-AMI erstellt werden. Die Amazon-EMR-Versionen 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 und höher enthalten einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“. Versionen mit einem niedrigeren Limit für geöffnete Dateien verursachen beim Senden des Spark-Jobs den Fehler „Zu viele offene Dateien“. In den betroffenen Versionen hat das Amazon-EMR-Standard-AMI eine Ulimit-StandardEinstellung von 4 096 für „Max. Anzahl geöffneter Dateien“, was unter dem Dateilimit von 65 536 im neuesten Amazon-Linux-2-AMI liegt. Die niedrigere Ulimit-Einstellung für „Max. Anzahl geöffneter Dateien“ führt dazu, dass der Spark-Job fehlschlägt, wenn der Spark-Treiber und der Executor versuchen, mehr als 4 096 Dateien zu öffnen. Um das Problem zu beheben, verfügt Amazon EMR über ein Bootstrap Action (BA)-Skript, das die Ulimit-Einstellung bei der Cluster-Erstellung anpasst.

Wenn Sie eine ältere Amazon-EMR-Version verwenden, für die dieses Problem nicht dauerhaft behoben ist, können Sie mit der folgenden Problemumgehung das Instance-Controller-Ulimit explizit auf maximal 65 536 Dateien festlegen.



## Explizit ein ulimit über die Befehlszeile setzen

1. Bearbeiten Sie `/etc/systemd/system/instance-controller.service`, um die folgenden Parameter zum Abschnitt Service hinzuzufügen.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Starten Sie neu InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

## Mithilfe der Bootstrap-Aktion (BA) ein Ulimit festlegen

Sie können auch ein Bootstrap-Aktionsskript (BA) verwenden, um das Ulimit für den Instance-Controller bei der Clustererstellung auf 65 536 Dateien zu konfigurieren.

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

- [EMR Notebooks](#)

Das Feature, mit der Sie Kernel und zusätzliche Python-Bibliotheken auf dem Cluster-Primärknoten installieren können, ist in EMR-Version 5.30.1 standardmäßig deaktiviert. Weitere Informationen zu diesem Feature finden Sie unter [Installation von Kernels und Python-Bibliotheken auf einem Cluster-Primärknoten](#).

Um das Feature zu aktivieren, gehen Sie wie folgt vor:

1. Stellen Sie sicher, dass die der Servicerolle für EMR Notebooks zugeordnete Berechtigungsrichtlinie die folgende Aktion zulässt:

```
elasticmapreduce:ListSteps
```

Weitere Informationen finden Sie unter [Servicerolle für EMR-Notebooks](#).

2. Verwenden Sie den AWS CLI , um einen Schritt auf dem Cluster auszuführen, der EMR Notebooks einrichtet, wie im folgenden Beispiel gezeigt. Ersetzen Sie *us-east-1* durch die Region, in der sich Ihr Cluster befindet. Weitere Informationen finden Sie unter [Hinzufügen von Schritten zu einem Cluster mithilfe AWS CLI](#).

```
aws emr add-steps --cluster-id MyClusterID --steps
  Type=CUSTOM_JAR,Name=EMRNotebooksSetup,ActionOnFailure=CONTINUE, Jar=s3://us-
east-1.elasticmapreduce/libs/script-runner/script-runner.jar, Args=["s3://
  awssupportdatasvcs.com/bootstrap-actions/EMRNotebooksSetup/emr-notebooks-setup.sh"]
```

- **Verwaltete Skalierung**

Verwaltete Skalierungsvorgänge auf Clustern der Versionen 5.30.0 und 5.30.1, ohne dass Presto installiert ist, können zu Anwendungsausfällen führen oder dazu führen, dass eine einheitliche Instance-Gruppe oder Instance-Flotte unverändert im Status ARRESTED bleibt, insbesondere wenn auf einen Herunterskalierungsvorgang schnell ein Skalierungsvorgang folgt.

Um dieses Problem zu umgehen, wählen Sie Presto als zu installierende Anwendung, wenn Sie einen Cluster mit den Amazon-EMR-Versionen 5.30.0 und 5.30.1 erstellen, auch wenn Ihr Auftrag Presto nicht benötigt.

- **Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung**

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in Amazon-EMR-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie dem Herunterskalieren oder der schrittweisen Übermittlung auftreten, nachdem der Cluster einige Zeit

lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

Workaround:

- SSH als hadoop-Benutzer für den führenden Primärknoten des EMR-Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den hadoop-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

#### Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

- Wenn Sie Spark mit der Formatierung des Hive-Partitionsstandorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den Amazon-EMR-Versionen 5.30.0 bis 5.36.0 und 6.2.0 bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:
  - Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.
  - Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist `s3://bucket/table/p=a` ein Präfix von `s3://bucket/table/p=a b`.
  - Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen UTF-8-Wert, der kleiner als das `/`-Zeichen (U+002F) ist. Beispielsweise fällt das Leerzeichen (U+0020), das in `s3://bucket/table/p=a b` zwischen `a` und `b` vorkommt, in diese Kategorie. Beachten Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: `!"#$%&'()*+,-.` Weitere Informationen finden Sie unter [UTF-8-Kodierungstabelle und Unicode-Zeichen](#).

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.

## Version 5.30.0

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.30.0. Änderungen beziehen sich auf Version 5.29.0.

Erste Version: 13. Mai 2020

Letzte Aktualisierung: 25. Juni 2020

### Upgrades

- Auf Version AWS SDK for Java 1.11.759 aktualisiert
- Amazon SageMaker Spark SDK wurde auf Version 1.3.0 aktualisiert
- EMR Record Server auf Version 1.6.0 aktualisiert
- Flink auf Version 1.10.0 aktualisiert
- Ganglia auf Version 3.7.2 aktualisiert
- HBase auf Version 1.4.13 aktualisiert
- Hudi auf Version 0.5.2-incubating aktualisiert
- Hue auf Version 4.6.0 aktualisiert
- Auf Version JupyterHub 1.1.0 aktualisiert
- Livy auf Version 0.7.0-incubating aktualisiert
- Oozie auf Version 5.2.0 aktualisiert
- Presto auf Version 0.232 aktualisiert
- Spark auf Version 2.4.5 aktualisiert
- Konnektoren und Treiber aktualisiert: Amazon Glue Connector 1.12.0; Amazon Kinesis Connector 3.5.0; EMR DynamoDB Connector 4.14.0

### Neue Features

- EMR Notebooks – Bei Verwendung mit EMR-Clustern, die mit 5.30.0 erstellt wurden, werden EMR-Notebooks-Kernel auf dem Cluster ausgeführt. Dies verbessert die Notebook-Leistung und

ermöglicht es Ihnen, Kernel zu installieren und anzupassen. Sie können Python-Bibliotheken auch auf dem Cluster-Primärknoten installieren. Weitere Informationen finden Sie unter [Installieren und Verwenden von Kernels und Bibliotheken](#) im Management Guide für EMR.

- **Verwaltete Skalierung** – Mit Amazon EMR ab Version 5.30.0 können Sie die verwaltete EMR-Skalierung aktivieren, um die Anzahl der Instances oder Einheiten in Ihrem Cluster basierend auf der Workload automatisch zu erhöhen oder zu verringern. Amazon EMR wertet Cluster-Metriken kontinuierlich aus, um Skalierungsentscheidungen zu treffen, die Ihre Cluster für Kosten und Geschwindigkeit optimieren. Weitere Informationen finden Sie unter [Skalieren von Clusterressourcen](#) im Verwaltungshandbuch für Amazon EMR.
- **In Amazon S3 gespeicherte Protokolldateien verschlüsseln** — Mit Amazon EMR Version 5.30.0 und höher können Sie in Amazon S3 gespeicherte Protokolldateien mit einem vom Kunden verwalteten Schlüssel verschlüsseln. AWS KMS Weitere Informationen finden Sie unter [Verschlüsselte gespeicherte Protokolldateien in Amazon S3](#) im Verwaltungshandbuch für Amazon EMR.
- **Unterstützung von Amazon Linux 2** – In EMR Version 5.30.0 und höher verwendet EMR das Betriebssystem Amazon Linux 2. Neue benutzerdefinierte AMIs (Amazon Machine Image) müssen auf dem Amazon-Linux-2-AMI basieren. Weitere Informationen finden Sie unter [Verwenden eines benutzerdefinierten AMI](#).
- **Ordnungsgemäßes Presto Auto Scaling** – EMR-Cluster mit 5.30.0 können mit einer Auto-Scaling-Zeitüberschreitung festgelegt werden, die Presto-Aufgaben Zeit zum Abschluss der Ausführung gibt, bevor ihr Knoten stillgelegt wird. Weitere Informationen finden Sie unter [Auto Scaling von Presto mit ordnungsgemäßer Stilllegung verwenden](#).
- **Flotteninstance-Erstellung mit neuer Zuweisungsstrategieoption** – Eine neue Zuweisungsstrategieoption ist in EMR-Version 5.12.1 und höher verfügbar. Sie bietet eine schnellere Cluster-Bereitstellung, eine genauere Spot-Zuweisung und weniger Unterbrechungen von Spot Instances. Aktualisierungen für nicht standardmäßige EMR-Service Rollen sind erforderlich. Sehen Sie unter [Konfigurieren von Instance-Flotten](#).
- **Befehle `sudo systemctl stop` und `sudo systemctl start`** – In EMR Version 5.30.0 und höher, die das Betriebssystem Amazon Linux 2 verwenden, verwendet EMR die Befehle `sudo systemctl stop` und `sudo systemctl start`, um Services neu zu starten. Weitere Informationen finden Sie unter [Wie starte ich einen Service in Amazon EMR?](#).

## Änderungen, Verbesserungen und behobene Probleme

- EMR Version 5.30.0 installiert Ganglia standardmäßig nicht. Sie können Ganglia explizit für die Installation auswählen, wenn Sie einen Cluster erstellen.
- Spark-Leistungsoptimierungen.
- Presto-Leistungsoptimierungen.
- Python 3 ist der Standard für Amazon-EMR-Version 5.30.0 und höher.
- Die standardmäßige verwaltete Sicherheitsgruppe für den Zugriff auf Services in privaten Subnetzen wurde mit neuen Regeln aktualisiert. Wenn Sie benutzerdefinierte Sicherheitsgruppe für den Servicezugriff verwenden, müssen Sie dieselben Regeln wie die standardmäßige verwaltete Sicherheitsgruppe einschließen. Weitere Informationen finden Sie unter [Amazon-EMR-verwaltete Sicherheitsgruppe für den Servicezugriff \(private Subnetze\)](#). Wenn Sie eine benutzerdefinierte Servicerolle für Amazon EMR verwenden, müssen Sie die Berechtigung zum `ec2:describeSecurityGroups` erteilen, damit EMR überprüfen kann, ob die Sicherheitsgruppen korrekt erstellt wurden. Wenn Sie `EMR_DefaultRole` verwenden, ist diese Berechtigung bereits in der standardmäßigen verwalteten Richtlinie enthalten.

## Bekannte Probleme

- Niedrigeres Limit für die „Max. Anzahl geöffneter Dateien“ auf älteren AL2-Versionen [in neueren Versionen behoben]. Amazon-EMR-Versionen: `emr-5.30.x`, `emr-5.31.0`, `emr-5.32.0`, `emr-6.0.0`, `emr-6.1.0` und `emr-6.2.0` basieren auf älteren Versionen von Amazon Linux 2 (AL2), die eine niedrigere `Ulimit`-Einstellung für „Max. Anzahl geöffneter Dateien“ haben, wenn Amazon-EMR-Cluster mit dem Standard-AMI erstellt werden. Die Amazon-EMR-Versionen `5.30.1`, `5.30.2`, `5.31.1`, `5.32.1`, `6.0.1`, `6.1.1`, `6.2.1`, `5.33.0`, `6.3.0` und höher enthalten einen dauerhaften Fix mit einer höheren Einstellung für „Max. Anzahl geöffneter Dateien“. Versionen mit einem niedrigeren Limit für geöffnete Dateien verursachen beim Senden des Spark-Jobs den Fehler „Zu viele offene Dateien“. In den betroffenen Versionen hat das Amazon-EMR-Standard-AMI eine `Ulimit`-Standardeinstellung von 4 096 für „Max. Anzahl geöffneter Dateien“, was unter dem Dateilimit von 65 536 im neuesten Amazon-Linux-2-AMI liegt. Die niedrigere `Ulimit`-Einstellung für „Max. Anzahl geöffneter Dateien“ führt dazu, dass der Spark-Job fehlschlägt, wenn der Spark-Treiber und der Executor versuchen, mehr als 4 096 Dateien zu öffnen. Um das Problem zu beheben, verfügt Amazon EMR über ein Bootstrap Action (BA)-Skript, das die `Ulimit`-Einstellung bei der Cluster-Erstellung anpasst.

Wenn Sie eine ältere Amazon-EMR-Version verwenden, für die dieses Problem nicht dauerhaft behoben ist, können Sie mit der folgenden Problemumgehung das Instance-Controller-Ulimit explizit auf maximal 65 536 Dateien festlegen.

Explizit ein ulimit über die Befehlszeile setzen

1. Bearbeiten Sie `/etc/systemd/system/instance-controller.service`, um die folgenden Parameter zum Abschnitt Service hinzuzufügen.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Starten Sie neu InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Mithilfe der Bootstrap-Aktion (BA) ein Ulimit festlegen

Sie können auch ein Bootstrap-Aktionsskript (BA) verwenden, um das Ulimit für den Instance-Controller bei der Clustererstellung auf 65 536 Dateien zu konfigurieren.

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
```

```
sudo systemctl daemon-reload
```

- **Verwaltete Skalierung**

Verwaltete Skalierungsvorgänge auf Clustern der Versionen 5.30.0 und 5.30.1, ohne dass Presto installiert ist, können zu Anwendungsausfällen führen oder dazu führen, dass eine einheitliche Instance-Gruppe oder Instance-Flotte unverändert im Status ARRESTED bleibt, insbesondere wenn auf einen Herunterskalierungsvorgang schnell ein Skalierungsvorgang folgt.

Um dieses Problem zu umgehen, wählen Sie Presto als zu installierende Anwendung, wenn Sie einen Cluster mit den Amazon-EMR-Versionen 5.30.0 und 5.30.1 erstellen, auch wenn Ihr Auftrag Presto nicht benötigt.

- **Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung**


Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in Amazon-EMR-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie dem Herunterskalieren oder der schrittweisen Übermittlung auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

Workaround:

- SSH als hadoop-Benutzer für den führenden Primärknoten des EMR-Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den hadoop-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

 **Note**

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen



konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

- Die Standard-Datenbank-Engine für Hue 4.6.0 ist SQLite, was zu Problemen führt, wenn Sie versuchen, Hue mit einer externen Datenbank zu verwenden. Um dieses Problem zu beheben, setzen Sie `engine` in Ihrer `hue.ini` Konfigurationsklassifizierung auf `mysql`. Dieses Problem wurde in Amazon-EMR-Version 5.30.1 behoben.
- Wenn Sie Spark mit der Formatierung des Hive-Partitionsstandorts verwenden, um Daten in Amazon S3 zu lesen, und Sie Spark auf den Amazon-EMR-Versionen 5.30.0 bis 5.36.0 und 6.2.0 bis 6.9.0 ausführen, kann ein Problem auftreten, das verhindert, dass Ihr Cluster Daten korrekt liest. Dies kann passieren, wenn Ihre Partitionen alle der folgenden Eigenschaften aufweisen:
  - Zwei oder mehr Partitionen werden aus derselben Tabelle gescannt.
  - Mindestens ein Partitionsverzeichnispfad ist ein Präfix für mindestens einen anderen Partitionsverzeichnispfad, z. B. ist `s3://bucket/table/p=a` ein Präfix von `s3://bucket/table/p=a b`.
  - Das erste Zeichen, das auf das Präfix im anderen Partitionsverzeichnis folgt, hat einen UTF-8-Wert, der kleiner als das `/`-Zeichen (U+002F) ist. Beispielsweise fällt das Leerzeichen (U+0020), das in `s3://bucket/table/p=a b` zwischen `a` und `b` vorkommt, in diese Kategorie. Beachten Sie, dass es 14 weitere Zeichen gibt, die keine Kontrollzeichen sind: `!"#$%&'()*+,-.` Weitere Informationen finden Sie unter [UTF-8-Kodierungstabelle und Unicode-Zeichen](#).

Um dieses Problem zu umgehen, stellen Sie die `spark.sql.sources.fastS3PartitionDiscovery.enabled`-Konfiguration auf `false` in der `spark-defaults`-Klassifizierung ein.

## Version 5.29.0

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.29.0. Änderungen beziehen sich auf Version 5.28.1.

Erste Version: 17. Januar 2020

### Upgrades

- Auf Version AWS SDK for Java 1.11.682 aktualisiert
- Upgrade von Hive auf Version 2.3.6
- Flink wurde auf Version 1.9.1 aktualisiert

- EmrFS wurde auf Version 2.38.0 aktualisiert
- EMR DynamoDB Connector wurde auf Version 4.13.0 aktualisiert

## Änderungen, Verbesserungen und behobene Probleme

- Spark
  - Spark-Leistungsoptimierungen.
- EMRFS
  - Management Guide aktualisiert die Standardeinstellungen von emrfs-site.xml für eine konsistente Ansicht.

## Bekannte Probleme

- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in Amazon-EMR-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie dem Herunterskalieren oder der schrittweisen Übermittlung auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

### Workaround:

- SSH als hadoop-Benutzer für den führenden Primärknoten des EMR-Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den hadoop-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

**Note**

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

## Version 5.28.1

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.28.1. Änderungen beziehen sich auf Version 5.28.0.

Erste Version: 10. Januar 2020

Änderungen, Verbesserungen und behobene Probleme

- Spark
  - Spark-Kompatibilitätsprobleme behoben.
- CloudWatch Metriken
  - Problem bei der Veröffentlichung von Amazon CloudWatch Metrics auf einem EMR-Cluster mit mehreren Primärknoten behoben.
- Deaktivierte Protokollnachricht
  - Falsche Protokollnachricht „... verwendet die alte Version (<4.5.8) des Apache-HTTP-Clients“ deaktiviert.

Bekannte Probleme

- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in Amazon-EMR-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie dem Herunterskalieren oder der schrittweisen Übermittlung auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

### Workaround:

- SSH als hadoop-Benutzer für den führenden Primärknoten des EMR-Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den hadoop-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

#### Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

## Version 5.28.0

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.28.0. Änderungen beziehen sich auf Version 5.27.0.

Erste Version: 12. November 2019

### Upgrades

- Upgrade von Flink auf Version 1.9.0
- Upgrade von Hive auf Version 2.3.6
- Upgrade von MXNet auf Version 1.5.1
- Upgrade von Phoenix auf Version 4.14.3
- Upgrade von Presto auf Version 0.227
- Upgrade von Zeppelin auf Version 0.8.2

## Neue Features

- [Apache Hudi](#) steht nun zur Installation unter Amazon EMR zur Verfügung, wenn Sie einen Cluster erstellen. Weitere Informationen finden Sie unter [Hudi](#).
- (25. November 2019) Sie können jetzt auf Wunsch mehrere Schritte gleichzeitig ausführen, um die Cluster-Nutzung zu verbessern und Kosten zu sparen. Außerdem können Sie sowohl anstehende als auch laufende Schritte stornieren. Weitere Informationen finden Sie unter [Arbeiten mit Schritten mithilfe der AWS CLI AND-Konsole](#).
- (3. Dezember 2019) Sie können jetzt EMR-Cluster erstellen und ausführen. AWS Outposts ermöglicht native AWS Dienste, Infrastrukturen und Betriebsmodelle in lokalen Einrichtungen. In AWS Outposts Umgebungen können Sie dieselben AWS APIs, Tools und Infrastrukturen verwenden, die Sie in der AWS Cloud verwenden. Weitere Informationen finden Sie unter [EMR-Cluster auf AWS Outposts](#).
- (11. März 2020) Ab Amazon EMR Version 5.28.0 können Sie Amazon EMR-Cluster in einem Local Zones-Subnetz als logische Erweiterung einer AWS Region, die AWS Local Zones unterstützt, erstellen und ausführen. Eine lokale Zone ermöglicht es, Amazon EMR-Funktionen und eine Untergruppe von AWS Diensten, wie Rechen- und Speicherdienste, näher an den Benutzern zu platzieren, wodurch der Zugriff auf lokal ausgeführte Anwendungen mit sehr geringer Latenz ermöglicht wird. Eine Liste der verfügbaren Local Zones finden Sie unter [AWS Local Zones](#). Informationen zum Zugriff auf verfügbare AWS Local Zones finden Sie unter [Regionen, Availability Zones und Local Zones](#).

Local Zones unterstützen derzeit keine Amazon EMR Notebooks und unterstützen keine direkten Verbindungen zu Amazon EMR über die Schnittstelle VPC-Endpunkt (AWS PrivateLink).

## Änderungen, Verbesserungen und behobene Probleme

- Erweiterter Anwendungssupport für Cluster mit hoher Verfügbarkeit
  - Weitere Informationen finden Sie unter [Unterstützte Anwendungen in einem EMR-Cluster mit mehreren Primärknoten](#) im Verwaltungshandbuch für Amazon EMR.
- Spark
  - Leistungsoptimierungen
- Hive
  - Leistungsoptimierungen
- Presto

- Leistungsoptimierungen

## Bekannte Probleme

- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in Amazon-EMR-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie dem Herunterskalieren oder der schrittweisen Übermittlung auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

### Workaround:

- SSH als hadoop-Benutzer für den führenden Primärknoten des EMR-Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den hadoop-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

### Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

## Veröffentlichung 5.27.0

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.27.0. Änderungen beziehen sich auf Version 5.26.0.

Erste Version: 23. September 2019

## Upgrades

- AWS SDK for Java 1.11.615
- Flink 1.8.1
- JupyterHub 1.0.0
- Spark 2.4.4
- Tensorflow 1,14.0
- Konnektoren und Treiber:
  - DynamoDB Connector 4.12.0

## Neue Features

- (24. Oktober 2019) Die folgenden neuen Features in EMR Notebooks sind in allen Amazon-EMR-Versionen verfügbar.
  - Sie können Git-Repositorys mit Ihren EMR-Notebooks verknüpfen, um Ihre Notebooks in einer versionskontrollierten Umgebung zu speichern. Sie können Code mit Kollegen teilen und vorhandene Jupyter Notebooks über Remote-Git-Repositorys wiederverwenden. Weitere Informationen finden Sie unter [Verknüpfen von Git-basierten Repositorys mit Amazon EMR Notebooks](#) im Verwaltungshandbuch für Amazon EMR.
  - Das [Hilfsprogramm nbdime](#) ist jetzt in EMR Notebooks verfügbar, um das Vergleichen und Zusammenführen von Notebooks zu vereinfachen.
  - EMR-Notebooks werden jetzt unterstützt JupyterLab. JupyterLab ist eine webbasierte interaktive Entwicklungsumgebung, die vollständig mit Jupyter-Notebooks kompatibel ist. Sie können jetzt wählen, ob Sie Ihr Notizbuch entweder im JupyterLab Jupyter-Notebook-Editor öffnen möchten.
- (30. Oktober 2019) Mit Amazon-EMR-Version 5.25.0 können Sie über die Seite Zusammenfassung des Clusters oder die Registerkarte Anwendungsverlauf in der Konsole eine Verbindung zu persistenten Spark History Servern herstellen. Anstatt einen Web-Proxy über eine SSH-Verbindung einzurichten, können Sie schnell auf die Benutzeroberfläche des Spark-History-Servers zugreifen, um Anwendungsmetriken einzusehen und auf relevante Protokolldateien für aktive und beendete Cluster zuzugreifen. Weitere Informationen finden Sie unter [Off-Cluster-Zugriff auf persistente Anwendungsbeneroberflächen](#) im Verwaltungshandbuch für Amazon EMR.

## Änderungen, Verbesserungen und behobene Probleme

- Amazon-EMR-Cluster mit mehreren Primärknoten
  - Sie können Flink auf einem Amazon-EMR-Cluster mit mehreren Primärknoten installieren und ausführen. Weitere Informationen finden Sie unter [Unterstützte Anwendungen und Features](#).
  - Sie können die transparente HDFS-Verschlüsselung auf einem Amazon-EMR-Cluster mit mehreren Primärknoten konfigurieren. Weitere Informationen finden Sie unter [Transparente HDFS-Verschlüsselung auf EMR-Clustern mit mehreren Primärknoten](#).
  - Sie können jetzt die Konfiguration von Anwendungen ändern, die auf einem Amazon-EMR-Cluster mit mehreren Primärknoten ausgeführt werden. Weitere Informationen finden Sie unter [Angaben einer Konfiguration für eine Instance-Gruppe in einem aktiven Cluster](#).
- Amazon-EMR-DynamoDB-Konnektor
  - Amazon-EMR-DynamoDB-Konnektor unterstützt jetzt die folgenden DynamoDB-Datentypen: boolean, list, map, item, null. Weitere Informationen finden Sie unter [Eine Hive-Tabelle zur Ausführung von Hive-Befehlen einrichten](#).

## Bekannte Probleme

- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in Amazon-EMR-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie dem Herunterskalieren oder der schrittweisen Übermittlung auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

### Workaround:

- SSH als hadoop-Benutzer für den führenden Primärknoten des EMR-Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den hadoop-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```



In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

#### Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

## Version 5.26.0

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.26.0. Änderungen beziehen sich auf Version 5.25.0.

Erste Version: 8. August 2019

Letzte Aktualisierung: 19. August 2019

### Upgrades

- AWS SDK for Java 1.11.595
- HBase 1.4.10
- Phoenix 4.14.2
- Konnektoren und Treiber:
  - DynamoDB Connector 4.11.0
  - MariaDB Connector 2.4.2
  - Amazon-Redshift-JDBC-Treiber 1.2.32.1056

### Neue Features

- (Beta) Mit Amazon EMR 5.26.0 können Sie einen Cluster starten, der in Lake Formation integriert ist. Diese Integration bietet einen detaillierten Zugriff auf Spaltenebene auf Datenbanken und Tabellen im Glue-Datenkatalog. AWS ermöglicht Lake Formation über ein Unternehmens-Identitätssystem eine verbundene einmalige Anmeldung bei EMR Notebooks oder Apache

Zeppelin. Weitere Informationen finden Sie unter [Integration von Amazon EMR mit AWS Lake Formation \(Beta\)](#).

- (19. August 2019) Amazon EMR Block Public Access ist jetzt in allen Amazon-EMR-Versionen verfügbar, die Sicherheitsgruppen unterstützen. Den öffentlichen Zugriff blockieren ist eine kontoweite Einstellung, die für jede Region gilt. AWS Block Public Access verhindert das Starten eines Clusters, wenn eine mit dem Cluster verknüpfte Sicherheitsgruppe über eine Regel verfügt, die eingehenden Datenverkehr von IPv4 0.0.0.0/0 oder IPv6 ::/0 (öffentlicher Zugriff) auf einem Port zulässt, sofern der Port nicht als Ausnahme festgelegt wurde. Port 22 ist standardmäßig eine Ausnahme. Weitere Informationen finden Sie unter [Verwenden von Amazon EMR Block Public Access](#) im Verwaltungshandbuch für Amazon EMR.

## Änderungen, Verbesserungen und behobene Probleme

- EMR Notebooks
  - Mit EMR 5.26.0 und höher unterstützt EMR Notebooks zusätzlich zu den Standard-Python-Bibliotheken auch Python-Bibliotheken für Notebooks. Sie können Bibliotheken für Notebooks vom Notebook-Editor aus installieren, ohne einen Cluster neu erstellen oder ein Notebook erneut an einen Cluster anhängen zu müssen. Notebook-bezogene Bibliotheken werden in einer virtuellen Python-Umgebung erstellt und gelten daher nur für die aktuelle Notebook-Sitzung. Auf diese Weise können Sie Abhängigkeiten zwischen Notebooks isolieren. Weitere Informationen finden Sie unter [Verwenden von Notebook Scoped Libraries](#) im Verwaltungshandbuch für Amazon EMR.
- EMRFS
  - Sie können ein ETag-Überprüfungsfeature (Beta) aktivieren, indem Sie `fs.s3.consistent.metadata.etag.verification.enabled` auf `true` einstellen. Mit diesem Feature verwendet EMRFS Amazon-S3-ETags, um zu überprüfen, ob gelesene Objekte die neueste verfügbare Version sind. Diese Funktion ist hilfreich für read-after-update Anwendungsfälle, in denen Dateien auf Amazon S3 überschrieben werden, während derselbe Name beibehalten wird. Diese ETag-Verifizierungsfunktion funktioniert derzeit nicht mit S3 Select. Weitere Informationen finden Sie unter [Konfigurieren der konsistenten Ansicht](#).
- Spark
  - Die folgenden Optimierungen sind jetzt standardmäßig aktiviert: dynamisches Bereinigen von Partitionen, DISTINCT vor INTERSECT, Verbesserungen bei der Inferenz von SQL-Planstatistiken für JOIN, gefolgt von DISTINCT-Abfragen, Reduzierung skalarer Unterabfragen,

optimierte Join-Neuordnung und Bloom-Filter-Join. Weitere Informationen finden Sie unter [Optimieren der Spark-Leistung](#).

- Die Codegenerierung für die gesamte Phase für Sort Merge Join wurde verbessert.
- Die Wiederverwendung von Abfragefragmenten und Unterabfragen wurde verbessert.
- Verbesserungen bei der Vorabzuweisung von Executoren beim Start von Spark.
- Bloom-Filter-Joins werden nicht mehr angewendet, wenn die kleinere Seite der Verknüpfung einen Broadcast-Hinweis enthält.
- Tez
  - Ein Problem mit Tez wurde behoben. Die Tez-Benutzeroberfläche funktioniert jetzt auf einem Amazon-EMR-Cluster mit mehreren Primärknoten.

### Bekannte Probleme

- Die verbesserten Funktionen zur Codegenerierung in der gesamten Phase für Sort Merge Join können den Speicherdruck erhöhen, wenn sie aktiviert sind. Diese Optimierung verbessert die Leistung, kann jedoch zu Wiederholungsversuchen oder Fehlschlägen führen, wenn der nicht so eingestellt `spark.yarn.executor.memoryOverheadFactor` ist, dass er genügend Speicher zur Verfügung stellt. Um dieses Feature zu deaktivieren, setzen Sie `spark.sql.sortMergeJoinExec.extendedCodegen.enabled` auf „false“.
- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in Amazon-EMR-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie dem Herunterskalieren oder der schrittweisen Übermittlung auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

### Workaround:

- SSH als hadoop-Benutzer für den führenden Primärknoten des EMR-Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den hadoop-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

#### Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

## Version 5.25.0

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.25.0. Änderungen beziehen sich auf Version 5.24.1.

Erste Version: 17. Juli 2019

Letzte Aktualisierung: 30. Oktober 2019

Amazon EMR 5.25.0

### Upgrades

- AWS SDK for Java 1.11.566
- Hive 2.3.5
- Presto 0.220
- Spark 2.4.3
- TensorFlow 1,13,1
- Tez 0.9.2
- Zookeeper 3.4.14

## Neue Features

- (30. Oktober 2019) Beginnend mit der Amazon-EMR-Version 5.25.0 können Sie über die SeiteZusammenfassung des Clusters oder die Registerkarte Anwendungsverlauf in der Konsole eine Verbindung zu persistenten Spark-History-Servern herstellen. Anstatt einen Web-Proxy über eine SSH-Verbindung einzurichten, können Sie schnell auf die Benutzeroberfläche des Spark-History-Servers zugreifen, um Anwendungsmetriken einzusehen und auf relevante Protokolldateien für aktive und beendete Cluster zuzugreifen. Weitere Informationen finden Sie unter [Off-Cluster-Zugriff auf persistente Anwendungsbenuzeroberflächen](#) im Verwaltungshandbuch für Amazon EMR.

## Änderungen, Verbesserungen und behobene Probleme

- Spark
  - Die Leistung einiger Verknüpfungen wurde verbessert, indem Bloom-Filter zum Vorfiltern von Eingaben verwendet wurden. Die Optimierung ist standardmäßig deaktiviert und kann aktiviert werden, indem der Spark-Konfigurationsparameter `spark.sql.bloomFilterJoin.enabled` auf `true` gesetzt wird.
  - Die Leistung bei der Gruppierung nach Spalten vom Typ Zeichenfolge wurde verbessert.
  - Die standardmäßige Speicher- und Kernkonfiguration des Spark-Executors von R4-Instance-Typen für Cluster ohne installiertes HBase wurde verbessert.
  - Es wurde ein früheres Problem mit dem Feature zum dynamischen Bereinigen von Partitionen behoben, bei dem sich die bereinigte Tabelle auf der linken Seite des Joins befinden musste.
  - Die Optimierung von DISTINCT vor INTERSECT wurde verbessert und gilt nun auch für weitere Fälle, in denen Aliase involviert sind.
  - Die Inferenz der SQL-Planstatistiken für JOIN, gefolgt von DISTINCT-Abfragen, wurde verbessert. Diese Verbesserung ist standardmäßig deaktiviert und kann aktiviert werden, indem der Spark-Konfigurationsparameter `spark.sql.statsImprovements.enabled` auf `true` gesetzt wird. Diese Optimierung ist für das Feature „Unterscheiden vor Überschneiden“ erforderlich und wird automatisch aktiviert, wenn `spark.sql.optimizer.distinctBeforeIntersect.enabled` auf `true` gesetzt ist.
  - Optimierte Verbindungsreihenfolge basierend auf Tabellengröße und Filtern. Diese Optimierung ist standardmäßig deaktiviert und kann aktiviert werden, indem der Spark-Konfigurationsparameter `spark.sql.optimizer.sizeBasedJoinReorder.enabled` auf `true` gesetzt wird.

Weitere Informationen finden Sie unter [Optimieren der Spark-Leistung](#).

- EMRFS

- Die EMRFS-Einstellung `fs.s3.buckets.create.enabled` ist jetzt standardmäßig deaktiviert. Bei Tests haben wir festgestellt, dass die Deaktivierung dieser Einstellung die Leistung verbessert und die unbeabsichtigte Erstellung von S3-Buckets verhindert. Wenn Ihre Anwendung auf dieser Funktionalität basiert, können Sie sie aktivieren, indem Sie die Eigenschaft `true` in der Konfigurationsklassifizierung von `fs.s3.buckets.create.enabled` auf `emrfs-site` setzen. Weitere Informationen finden Sie unter [Angeben einer Konfiguration beim Erstellen eines Clusters](#).
- Verbesserungen der lokalen Festplattenverschlüsselung und der S3-Verschlüsselung in Sicherheitskonfigurationen (5. August 2019)
  - Die Amazon-S3-Verschlüsselungseinstellungen wurden im Sicherheitskonfigurations-Setup von den Einstellungen für die lokale Festplattenverschlüsselung getrennt.
  - Es wurde eine Option hinzugefügt, um die EBS-Verschlüsselung mit Version 5.24.0 und höher zu aktivieren. Bei Auswahl dieser Option wird zusätzlich zu den Speichervolumen auch das Root-Geräte-Volumen verschlüsselt. Frühere Versionen erforderten die Verwendung eines benutzerdefinierten AMI, um das Root-Geräte-Volumen zu verschlüsseln.
  - Weitere Informationen erhalten Sie unter [Verschlüsselungsoptionen](#) im Verwaltungshandbuch für Amazon EMR.

## Bekannte Probleme

- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in Amazon-EMR-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie dem Herunterskalieren oder der schrittweisen Übermittlung auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

### Workaround:

- SSH als `hadoop`-Benutzer für den führenden Primärknoten des EMR-Clusters mit mehreren Primärknoten.

- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den hadoop-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

#### Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

## Version 5.24.1

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.24.1. Änderungen beziehen sich auf Version 5.24.0.

Erste Version: 26. Juni 2019

Änderungen, Verbesserungen und behobene Probleme

- Das standardmäßige Amazon Linux AMI für Amazon EMR wurde aktualisiert, sodass es wichtige Sicherheitsupdates für den Linux-Kernel enthält, darunter das „TCP SACK Denial of Service Issue“ ([AWS-2019-005](#)).

Bekannte Probleme

- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in Amazon-EMR-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie dem Herunterskalieren oder der schrittweisen Übermittlung auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische

als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

Workaround:

- SSH als hadoop-Benutzer für den führenden Primärknoten des EMR-Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den hadoop-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

#### Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

## Version 5.24.0

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.24.0. Änderungen beziehen sich auf Version 5.23.0.

Erste Version: 11. Juni 2019

Letzte Aktualisierung: 5. August 2019

### Upgrades

- Flink 1.8.0
- Hue 4.4.0
- JupyterHub 0.9.6
- Livy 0.6.0
- MxNet 1.4.0



- Presto 0.219
- Spark 2.4.2
- AWS SDK for Java 1,11.546
- Konnektoren und Treiber:
  - DynamoDB Connector 4.9.0
  - MariaDB Connector 2.4.1
  - Amazon-Redshift-JDBC-Treiber, Version 1.2.27.1051

## Änderungen, Verbesserungen und behobene Probleme

- Spark
  - Es wurde eine Optimierung für das dynamische Bereinigen von Partitionen hinzugefügt. Die Optimierung ist standardmäßig deaktiviert. Um es zu aktivieren, setzen Sie den Spark-Konfigurationsparameter `spark.sql.dynamicPartitionPruning.enabled` auf `true`.
  - Die Leistung von INTERSECT-Anfragen wurde verbessert. Diese Optimierung ist standardmäßig deaktiviert. Um es zu aktivieren, setzen Sie den Spark-Konfigurationsparameter `spark.sql.optimizer.distinctBeforeIntersect.enabled` auf `true`.
  - Es wurde eine Optimierung hinzugefügt, um skalare Unterabfragen mit Aggregaten, die dieselbe Beziehung verwenden, zu vereinfachen. Die Optimierung ist standardmäßig deaktiviert. Um es zu aktivieren, setzen Sie den Spark-Konfigurationsparameter `spark.sql.optimizer.flattenScalarSubqueriesWithAggregates.enabled` auf `true`.
  - Die Codegenerierung für die gesamte Phase wurde verbessert.

Weitere Informationen finden Sie unter [Optimieren der Spark-Leistung](#).

- Verbesserungen der lokalen Festplattenverschlüsselung und der S3-Verschlüsselung in Sicherheitskonfigurationen (5. August 2019)
  - Die Amazon-S3-Verschlüsselungseinstellungen wurden im Sicherheitskonfigurations-Setup von den Einstellungen für die lokale Festplattenverschlüsselung getrennt.
  - Es wurde eine Option hinzugefügt, um die EBS-Verschlüsselung zu aktivieren. Bei Auswahl dieser Option wird zusätzlich zu den Speichervolumen auch das Root-Geräte-Volumen verschlüsselt. Frühere Versionen erforderten die Verwendung eines benutzerdefinierten AMI, um das Root-Geräte-Volumen zu verschlüsseln.

- Weitere Informationen erhalten Sie unter [Verschlüsselungsoptionen](#) im Verwaltungshandbuch für Amazon EMR.

## Bekannte Probleme

- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in Amazon-EMR-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie dem Herunterskalieren oder der schrittweisen Übermittlung auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

### Workaround:

- SSH als `hadoop`-Benutzer für den führenden Primärknoten des EMR-Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den `hadoop`-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

### Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

## Version 5.23.0

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.23.0. Änderungen beziehen sich auf Version 5.22.0.

Erste Version: 01. April 2019

Letzte Aktualisierung: 30. April 2019

### Upgrades

- AWS SDK for Java 1,11.519

### Neue Features

- (30. April 2019) Mit Amazon EMR 5.23.0 und höher können Sie einen Cluster mit drei primären Knoten starten, um die Hochverfügbarkeit von Anwendungen wie YARN Resource Manager, HDFS, Spark NameNode, Hive und Ganglia zu unterstützen. Der Primärknoten ist mit diesem Feature keine potenzielle einzelne Fehlerquelle mehr. Wenn ein Primärknoten ausfällt, führt Amazon EMR automatisch einen Failover zu einem Standby-Primärknoten aus und ersetzt den ausgefallenen Primärknoten durch einen neuen Primärknoten mit der gleichen Konfiguration und den gleichen Bootstrap-Aktionen. Weitere Informationen finden Sie unter [Primärknoten planen und konfigurieren](#).

### Bekannte Probleme

- Tez UI (In Amazon-EMR-Version 5.26.0 behoben)

Die Tez-Benutzeroberfläche funktioniert nicht auf einem EMR-Cluster mit mehreren Primärknoten.

- Hue (in Amazon-EMR-Version 5.24.0 behoben)
  - Hue, das auf Amazon EMR läuft, unterstützt Solr nicht. Ab Amazon-EMR-Version 5.20.0 führt ein Problem mit der Fehlkonfiguration dazu, dass Solr aktiviert wird und eine harmlose Fehlermeldung ähnlich der folgenden angezeigt wird:

```
Solr server could not be contacted properly:
HTTPConnectionPool('host=ip-xx-xx-xx-xx.ec2.internal',
port=1978): Max retries exceeded with url: /solr/admin/info/
system?user.name=hue&doAs=administrator&wt=json (Caused by
```

```
NewConnectionError(': Failed to establish a new connection: [Errno 111]
Connection refused',))
```

So wird verhindert, dass die Solr-Fehlermeldung angezeigt wird:

1. Stellen Sie über SSH eine Verbindung zur Befehlszeile des Primärknotens her.
2. Verwenden Sie einen Texteditor zum Öffnen der `hue.ini`-Datei. Beispielsweise:

```
sudo vim /etc/hue/conf/hue.ini
```

3. Suchen Sie nach dem Begriff `appblacklist` und ändern Sie die Zeile wie folgt:

```
appblacklist = search
```

4. Speichern Sie Ihre Änderungen und starten Sie Hue wie im folgenden Beispiel gezeigt:

```
sudo stop hue; sudo start hue
```

- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung


Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in Amazon-EMR-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie dem Herunterskalieren oder der schrittweisen Übermittlung auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

Workaround:

- SSH als `hadoop`-Benutzer für den führenden Primärknoten des EMR-Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den `hadoop`-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```


In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

 Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

## Version 5.22.0

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.22.0. Änderungen beziehen sich auf Version 5.21.0.

 Important

Ab Amazon EMR Version 5.22.0 verwendet Amazon EMR AWS Signature Version 4 ausschließlich zur Authentifizierung von Anfragen an Amazon S3. Frühere Amazon EMR-Versionen verwenden in einigen Fällen AWS Signature Version 2, sofern in den Versionshinweisen nicht angegeben ist, dass ausschließlich Signature Version 4 verwendet wird. Weitere Informationen finden Sie unter [Authentifizieren von Anfragen \(AWS Signature Version 4\)](#) und [Authentifizieren von Anfragen \(AWS Signature Version 2\)](#) im Amazon Simple Storage Service Developer Guide.

Erste Version: 20. März 2019

### Upgrades

- Flink 1.7.1
- HBase 1.4.9
- Oozie 5.1.0
- Phoenix 4.14.1
- Zeppelin 0.8.1
- Konnektoren und Treiber:
  - DynamoDB Connector 4.8.0

- MariaDB Connector 2.2.6
- Amazon-Redshift-JDBC-Treiber, Version 1.2.20.1043

## Neue Features

- Die EBS-Standardkonfiguration für EC2-Instance-Typen mit reinem EBS-Speicher wurde geändert. Beim Erstellen eines Clusters mit Amazon-EMR-Version ab 5.22.0 steigt die Größe des standardmäßig zugewiesenen EBS-Speichers basierend auf der Größe der Instance. Darüber hinaus haben wir den erhöhten Speicher auf mehrere Volumes aufgeteilt, was zu einer höheren IOPS-Leistung führt. Wenn Sie eine andere EBS-Instance-Speicherkonfiguration verwenden möchten, können Sie diese beim Erstellen eines EMR-Clusters bzw. beim Hinzufügen von Knoten zu einem Cluster angeben. Weitere Informationen zur Speichermenge und Anzahl der standardmäßig zugewiesenen Volumes für jeden Instance-Typ finden Sie unter [Standard-EBS-Speicher für Instances](#) im Verwaltungshandbuch für Amazon EMR.

## Änderungen, Verbesserungen und behobene Probleme

- Spark
  - Es wurde eine neue Konfigurationseigenschaft für Spark auf YARN `spark.yarn.executor.memoryOverheadFactor` eingeführt. Der Wert dieser Eigenschaft ist ein Skalierungsfaktor, der den Wert des Speicher-Overheads auf einen Prozentsatz des Executor-Speichers festlegt, mit einem Minimum von 384 MB. Wenn der Speicher-Overhead explizit mit `spark.yarn.executor.memoryOverhead` festgelegt wird, hat diese Eigenschaft keine Auswirkung. Der Standardwert ist `0.1875`, was 18,75 % entspricht. Dieser Standard für Amazon EMR lässt mehr Platz in YARN-Containern für den Executor-Speicheraufwand übrig als der intern von Spark festgelegte Standard von 10 %. Der Amazon-EMR-Standard von 18,75 % ergab empirisch weniger speicherbedingte Ausfälle in TPC-DS-Benchmarks.
  - [SPARK-26316](#) wurde zurückportiert, um die Leistung zu verbessern.
- In den Amazon-EMR-Versionen 5.19.0, 5.20.0 und 5.21.0 werden YARN-Knotenbezeichnungen in einem HDFS-Verzeichnis gespeichert. In einigen Situationen führt dies zu Verzögerungen beim Startup des Core-Knotens und dann zu einem Cluster-Timeout und einem Startfehler. Ab Amazon EMR 5.22.0 ist dieses Problem behoben. YARN-Knotenbezeichnungen werden auf der lokalen Festplatte jedes Clusterknotens gespeichert, wodurch Abhängigkeiten von HDFS vermieden werden.

## Bekannte Probleme

- Hue (in Amazon-EMR-Version 5.24.0 behoben)
  - Hue, das auf Amazon EMR läuft, unterstützt Solr nicht. Ab Amazon-EMR-Version 5.20.0 führt ein Problem mit der Fehlkonfiguration dazu, dass Solr aktiviert wird und eine harmlose Fehlermeldung ähnlich der folgenden angezeigt wird:

```
Solr server could not be contacted properly:
HTTPConnectionPool('host=ip-xx-xx-xx-xx.ec2.internal',
port=1978): Max retries exceeded with url: /solr/admin/info/
system?user.name=hue&doAs=administrator&wt=json (Caused by
NewConnectionError(': Failed to establish a new connection: [Errno 111]
Connection refused',))
```

So wird verhindert, dass die Solr-Fehlermeldung angezeigt wird:

1. Stellen Sie über SSH eine Verbindung zur Befehlszeile des Primärknotens her.
2. Verwenden Sie einen Texteditor zum Öffnen der `hue.ini`-Datei. Beispielsweise:

```
sudo vim /etc/hue/conf/hue.ini
```

3. Suchen Sie nach dem Begriff `appblacklist` und ändern Sie die Zeile wie folgt:

```
appblacklist = search
```

4. Speichern Sie Ihre Änderungen und starten Sie Hue wie im folgenden Beispiel gezeigt:

```
sudo stop hue; sudo start hue
```

- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in Amazon-EMR-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie dem Herunterskalieren oder der schrittweisen Übermittlung auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

Workaround:

- SSH als hadoop-Benutzer für den führenden Primärknoten des EMR-Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den hadoop-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

#### Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

## Version 5.21.1

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.21.1. Änderungen beziehen sich auf Version 5.21.0.

Erste Version: 18. Juli 2019

Änderungen, Verbesserungen und behobene Probleme

- Das standardmäßige Amazon Linux AMI für Amazon EMR wurde aktualisiert, sodass es wichtige Sicherheitsupdates für den Linux-Kernel enthält, darunter das „TCP SACK Denial of Service Issue“ ([AWS-2019-005](#)).

Bekannte Probleme

- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in Amazon-EMR-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie dem Herunterskalieren oder der schrittweisen Übermittlung auftreten, nachdem der Cluster einige Zeit



lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

Workaround:

- SSH als hadoop-Benutzer für den führenden Primärknoten des EMR-Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den hadoop-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

#### Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

## Version 5.21.0

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.21.0. Änderungen beziehen sich auf Version 5.20.0.

Veröffentlichungsdatum: 18. Februar 2019

Letzte Aktualisierung: 3. April 2019

### Upgrades

- Flink 1.7.0
- Presto 0.215
- AWS SDK for Java 1.11.479

## Neue Features

- (3. April 2019) Ab Amazon-EMR-Version 5.21.0 können Sie Cluster-Konfigurationen überschreiben und zusätzliche Konfigurationsklassifikationen für jede Instance-Gruppe in einem ausgeführten Cluster angeben. Dazu verwenden Sie die Amazon EMR-Konsole, das AWS Command Line Interface (AWS CLI) oder das AWS SDK. Weitere Informationen finden Sie unter [Angeben einer Konfiguration für eine Instance-Gruppe in einem aktiven Cluster](#).

## Änderungen, Verbesserungen und behobene Probleme

- Zeppelin
  - [ZEPPELIN-3878](#) rückportiert.

## Bekannte Probleme

- Hue (in Amazon-EMR-Version 5.24.0 behoben)
  - Hue, das auf Amazon EMR läuft, unterstützt Solr nicht. Ab Amazon-EMR-Version 5.20.0 führt ein Problem mit der Fehlkonfiguration dazu, dass Solr aktiviert wird und eine harmlose Fehlermeldung ähnlich der folgenden angezeigt wird:

```
Solr server could not be contacted properly:
HTTPConnectionPool('host=ip-xx-xx-xx-xx.ec2.internal',
port=1978): Max retries exceeded with url: /solr/admin/info/
system?user.name=hue&doAs=administrator&wt=json (Caused by
NewConnectionError(': Failed to establish a new connection: [Errno 111]
Connection refused',))
```

So wird verhindert, dass die Solr-Fehlermeldung angezeigt wird:

1. Stellen Sie über SSH eine Verbindung zur Befehlszeile des Primärknotens her.
2. Verwenden Sie einen Texteditor zum Öffnen der hue.ini-Datei. Beispielsweise:

```
sudo vim /etc/hue/conf/hue.ini
```

3. Suchen Sie nach dem Begriff `appblacklist` und ändern Sie die Zeile wie folgt:

```
appblacklist = search
```

4. Speichern Sie Ihre Änderungen und starten Sie Hue wie im folgenden Beispiel gezeigt:

```
sudo stop hue; sudo start hue
```

- Tez
  - Dieses Problem wurde in Amazon EMR 5.22.0 behoben.

Wenn Sie über eine SSH-Verbindung zum Primärknoten des Clusters eine Verbindung zur Tez-Benutzeroberfläche unter `http://MasterDNS:8080/tez-ui` herstellen, wird der Fehler „Adapterbetrieb fehlgeschlagen – Timeline-Server (ATS) ist nicht erreichbar“ angezeigt. Entweder ist er ausgefallen oder CORS ist nicht aktiviert“ wird angezeigt, oder Aufgaben zeigen unerwartet „N/A“ an.

Dies wird dadurch verursacht, dass die Tez-Benutzeroberfläche Anfragen an den YARN Timeline Server stellt und `localhost` nicht den Hostnamen des Primärknotens verwendet. Um dieses Problem zu umgehen, steht ein Skript zur Verfügung, das als Bootstrap-Aktion oder als Bootstrap-Schritt ausgeführt werden kann. Das Skript aktualisiert den Hostnamen in der `configs.env` Tez-Datei. Weitere Informationen und den Speicherort des Skripts finden Sie in den [Bootstrap-Anweisungen](#).

- In den Amazon-EMR-Versionen 5.19.0, 5.20.0 und 5.21.0 werden YARN-Knotenbezeichnungen in einem HDFS-Verzeichnis gespeichert. In einigen Situationen führt dies zu Verzögerungen beim Startup des Core-Knotens und dann zu einem Cluster-Timeout und einem Startfehler. Ab Amazon EMR 5.22.0 ist dieses Problem behoben. YARN-Knotenbezeichnungen werden auf der lokalen Festplatte jedes Clusterknotens gespeichert, wodurch Abhängigkeiten von HDFS vermieden werden.
- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in Amazon-EMR-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie dem Herunterskalieren oder der schrittweisen Übermittlung auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

Workaround:

- SSH als `hadoop`-Benutzer für den führenden Primärknoten des EMR-Clusters mit mehreren Primärknoten.

- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den hadoop-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

#### Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

## Version 5.20.0

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.20.0. Änderungen beziehen sich auf Version 5.19.0.

Veröffentlichungsdatum: 18. Dezember 2018

Letzte Aktualisierung: 22. Januar 2019

### Upgrades

- Flink 1.6.2
- HBase 1.4.8
- Hive 2.3.4
- Hue 4.3.0
- MXNet 1.3.1
- Presto 0.214
- Spark 2.4.0
- TensorFlow 1.12.0
- Tez 0.9.1

- AWS SDK for Java 1.11.461

## Neue Features

- (22. Januar 2019) Kerberos in Amazon EMR wurde verbessert und unterstützt nun die Authentifizierung von Prinzipalen von einem externen KDC. Dadurch wird die Prinzipalverwaltung zentralisiert, da mehrere Cluster einen einzelnen externen KDC gemeinsam nutzen können. Darüber hinaus kann der externe KDC eine bereichsübergreifende Vertrauensstellung mit einer Active-Directory-Domain haben. Auf diese Weise können alle Cluster Prinzipale von Active Directory authentifizieren. Weitere Informationen finden Sie unter [Verwendung der Kerberos-Authentifizierung](#) im Verwaltungshandbuch für Amazon EMR.

## Änderungen, Verbesserungen und behobene Probleme

- Standard-Amazon-Linux-AMI für Amazon EMR
  - Das Python3-Paket wurde von Python 3.4 auf 3.6 aktualisiert.
- Der EMRFS-S3-optimierte Committer
  - Der für EMRFS S3 optimierte Committer ist jetzt standardmäßig aktiviert, was die Schreibleistung verbessert. Weitere Informationen finden Sie unter [EMRFS-S3-optimierte Committer verwenden](#).
- Hive
  - [HIVE-16686](#) wurde rückportiert.
- Mit Spark und Hive Glue
  - In EMR 5.20.0 oder höher wird das parallel Partitionsbereinigen automatisch für Spark und Hive aktiviert, wenn AWS Glue Data Catalog als Metastore verwendet wird. Diese Änderung reduziert die Zeit für die Abfrageplanung erheblich, da mehrere Anfragen parallel ausgeführt werden, um Partitionen abzurufen. Die Gesamtzahl der Segmente, die gleichzeitig ausgeführt werden können, liegt zwischen 1 und 10. Der Standardwert ist 5, was eine empfohlene Einstellung ist. Sie können dies ändern, indem Sie die Eigenschaft `aws.glue.partition.num.segments` in der `hive-site`-Konfigurationsklassifizierung angeben. Wenn eine Drosselung auftritt, können Sie das Feature ausschalten, indem Sie den Wert auf 1 ändern. Weitere Informationen finden Sie unter [AWS -Glue-Segmentstruktur](#).

## Bekannte Probleme

- Hue (in Amazon-EMR-Version 5.24.0 behoben)

- Hue, das auf Amazon EMR läuft, unterstützt Solr nicht. Ab Amazon-EMR-Version 5.20.0 führt ein Problem mit der Fehlkonfiguration dazu, dass Solr aktiviert wird und eine harmlose Fehlermeldung ähnlich der folgenden angezeigt wird:

```
Solr server could not be contacted properly:
HTTPConnectionPool('host=ip-xx-xx-xx-xx.ec2.internal',
port=1978): Max retries exceeded with url: /solr/admin/info/
system?user.name=hue&doAs=administrator&wt=json (Caused by
NewConnectionError(': Failed to establish a new connection: [Errno 111]
Connection refused',))
```

So wird verhindert, dass die Solr-Fehlermeldung angezeigt wird:

1. Stellen Sie über SSH eine Verbindung zur Befehlszeile des Primärknotens her.
2. Verwenden Sie einen Texteditor zum Öffnen der `hue.ini`-Datei. Beispielsweise:

```
sudo vim /etc/hue/conf/hue.ini
```

3. Suchen Sie nach dem Begriff `appblacklist` und ändern Sie die Zeile wie folgt:

```
appblacklist = search
```

4. Speichern Sie Ihre Änderungen und starten Sie Hue wie im folgenden Beispiel gezeigt:

```
sudo stop hue; sudo start hue
```

- Tez
  - Dieses Problem wurde in Amazon EMR 5.22.0 behoben.

Wenn Sie über eine SSH-Verbindung zum Primärknoten des Clusters eine Verbindung zur Tez-Benutzeroberfläche unter `http://MasterDNS:8080/tez-ui` herstellen, wird der Fehler „Adapterbetrieb fehlgeschlagen – Timeline-Server (ATS) ist nicht erreichbar“ angezeigt. Entweder ist er ausgefallen oder CORS ist nicht aktiviert“ wird angezeigt, oder Aufgaben zeigen unerwartet „N/A“ an.

Dies wird dadurch verursacht, dass die Tez-Benutzeroberfläche Anfragen an den YARN Timeline Server stellt und `localhost` nicht den Hostnamen des Primärknotens verwendet. Um dieses Problem zu umgehen, steht ein Skript zur Verfügung, das als Bootstrap-Aktion oder als Bootstrap-Schritt ausgeführt werden kann. Das Skript aktualisiert den Hostnamen in der

`configs.env` Tez-Datei. Weitere Informationen und den Speicherort des Skripts finden Sie in den [Bootstrap-Anweisungen](#).

- In den Amazon-EMR-Versionen 5.19.0, 5.20.0 und 5.21.0 werden YARN-Knotenbezeichnungen in einem HDFS-Verzeichnis gespeichert. In einigen Situationen führt dies zu Verzögerungen beim Startup des Core-Knotens und dann zu einem Cluster-Timeout und einem Startfehler. Ab Amazon EMR 5.22.0 ist dieses Problem behoben. YARN-Knotenbezeichnungen werden auf der lokalen Festplatte jedes Clusterknotens gespeichert, wodurch Abhängigkeiten von HDFS vermieden werden.
- Bekanntes Problem in Clustern mit mehreren Primärknoten und Kerberos-Authentifizierung

Wenn Sie Cluster mit mehreren Primärknoten und Kerberos-Authentifizierung in Amazon-EMR-Versionen 5.20.0 und höher ausführen, können Probleme mit Cluster-Vorgängen wie dem Herunterskalieren oder der schrittweisen Übermittlung auftreten, nachdem der Cluster einige Zeit lang ausgeführt wurde. Der Zeitraum hängt von der Gültigkeitsdauer des Kerberos-Tickets ab, die Sie definiert haben. Das Herunterskalierungs-Problem wirkt sich sowohl auf automatische als auch auf explizite Herunterskalierungs-Anfragen aus, die Sie eingereicht haben. Weitere Clustervorgänge können ebenfalls beeinträchtigt werden.

Workaround:

- SSH als `hadoop`-Benutzer für den führenden Primärknoten des EMR-Clusters mit mehreren Primärknoten.
- Führen Sie den folgenden Befehl aus, um das Kerberos-Ticket für den `hadoop`-Benutzer zu erneuern.

```
kinit -kt <keytab_file> <principal>
```

In der Regel befindet sich die Keytab-Datei unter `/etc/hadoop.keytab` und der Prinzipal hat das Format von `hadoop/<hostname>@<REALM>`.

#### Note

Diese Problemumgehung gilt für den Zeitraum, in dem das Kerberos-Ticket gültig ist. Diese Dauer beträgt standardmäßig 10 Stunden, kann aber anhand Ihrer Kerberos-Einstellungen konfiguriert werden. Sie müssen den obigen Befehl erneut ausführen, sobald das Kerberos-Ticket abgelaufen ist.

## Version 5.19.0

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.19.0. Änderungen beziehen sich auf Version 5.18.0.

Erste Version: 7. November 2018

Letzte Aktualisierung: 19. November 2018

### Upgrades

- Hadoop 2.8.5
- Flink 1.6.1
- JupyterHub 0.9.4
- MXNet 1.3.0
- Presto 0.212
- TensorFlow 1.11.0
- Zookeeper 3.4.13
- AWS SDK for Java 1.11.433

### Neue Features

- (19. November 2018) EMR Notebooks ist eine verwaltete Umgebung, die auf Jupyter Notebook basiert. Es unterstützt Spark-Magic-Kernel für PySpark, Spark SQL, Spark R und Scala. EMR Notebooks können mit Clustern verwendet werden, die mit Amazon-EMR-Version 5.18.0 und höher erstellt wurden. Weitere Informationen finden Sie unter [Nutzung von EMR Notebooks](#) im Verwaltungshandbuch für Amazon EMR.
- Der S3-optimierte EMRFS-Committer ist beim Schreiben von Parquet-Dateien mit Spark und EMRFS verfügbar. Dieser Committer verbessert die Schreibleistung. Weitere Informationen finden Sie unter [EMRFS-S3-optimierte Committer verwenden](#).

### Änderungen, Verbesserungen und behobene Probleme

- YARN
  - Die Logik, die den Anwendungsmasterprozess auf die Ausführung auf Core-Knoten beschränkt, wurde geändert. Diese Funktionalität verwendet jetzt das Feature und Eigenschaften von YARN-Knotenbeschriftungen in den Klassifizierungen `yarn-site` und `der capacity-scheduler`



Konfiguration. Weitere Informationen finden Sie unter <https://docs.aws.amazon.com/emr/latest/ManagementGuide/emr-plan-instances-guidelines.html#emr-plan-spot-YARN>.

- Standard-Amazon-Linux-AMI für Amazon EMR
  - `ruby18`, `php56` und `gcc48` sind nicht mehr standardmäßig installiert. Diese können auf Wunsch mit `yum` installiert werden.
  - Das `aws-sdk` Ruby Gem ist standardmäßig nicht mehr installiert. Es kann, falls gewünscht, mit `gem install aws-sdk` installiert werden. Bestimmte Komponenten können ebenfalls installiert werden. z. B. `gem install aws-sdk-s3`.

## Bekannte Probleme

- EMR Notebooks – Unter bestimmten Umständen kann es vorkommen, dass der Notebook-Editor keine Verbindung zum Cluster herstellen kann, wenn mehrere Notebook-Editoren geöffnet sind. Löschen Sie in diesem Fall die Browser-Cookies und öffnen Sie die Notebook-Editoren erneut.
- CloudWatch ContainerPending Metrische und automatische Skalierung — (In 5.20.0 behoben) Amazon EMR gibt möglicherweise einen negativen Wert für `ContainerPending` aus. Wenn `ContainerPending` in einer Auto-Scaling-Regel verwendet wird, verhält sich das Auto Scaling nicht wie erwartet. Vermeiden Sie die Verwendung von `ContainerPending` mit automatischer Skalierung.
- In den Amazon-EMR-Versionen 5.19.0, 5.20.0 und 5.21.0 werden YARN-Knotenbezeichnungen in einem HDFS-Verzeichnis gespeichert. In einigen Situationen führt dies zu Verzögerungen beim Startup des Core-Knotens und dann zu einem Cluster-Timeout und einem Startfehler. Ab Amazon EMR 5.22.0 ist dieses Problem behoben. YARN-Knotenbezeichnungen werden auf der lokalen Festplatte jedes Clusterknotens gespeichert, wodurch Abhängigkeiten von HDFS vermieden werden.

## Version 5.18.0

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.18.0. Änderungen beziehen sich auf Version 5.17.0.

Veröffentlichungsdatum: 24. Oktober 2018

### Upgrades

- Flink 1.6.0

- HBase 1.4.7
- Presto 0.210
- Spark 2.3.2
- Zeppelin 0.8.0

## Neue Features

- Ab Amazon EMR 5.18.0 können Sie das Amazon-EMR-Artefakt-Repository verwenden, um Ihren Auftragscode anhand der genauen Versionen von Bibliotheken und Abhängigkeiten zu erstellen, die mit bestimmten Amazon-EMR-Versionen verfügbar sind. Weitere Informationen finden Sie unter [Überprüfen von Abhängigkeiten mithilfe des Amazon-EMR-Artefakt-Repositorys](#).

## Änderungen, Verbesserungen und behobene Probleme

- Hive
  - Unterstützung für S3 Select hinzugefügt. Weitere Informationen finden Sie unter [Verwendung von S3 Select mit Hive zur Verbesserung der Leistung](#).
- Presto
  - Unterstützung für [S3 Select](#) Pushdown hinzugefügt. Weitere Informationen finden Sie unter [S3-Select-Pushdown mit Presto zur Verbesserung der Leistung verwenden](#).
- Spark
  - Die standardmäßige log4j-Konfiguration für Spark wurde dahingehend geändert, dass Container-Logs für Spark-Streaming-Aufträge stündlich angezeigt werden. Dadurch wird verhindert, dass Logs für Spark-Streaming-Aufträge mit langer Laufzeit gelöscht werden.

## Version 5.17.1

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.17.1. Änderungen beziehen sich auf Version 5.17.0.

Erste Version: 18. Juli 2019

## Änderungen, Verbesserungen und behobene Probleme

- Das standardmäßige Amazon Linux AMI für Amazon EMR wurde aktualisiert, sodass es wichtige Sicherheitsupdates für den Linux-Kernel enthält, darunter das „TCP SACK Denial of Service Issue“ ([AWS-2019-005](#)).

## Version 5.17.0

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.17.0. Änderungen beziehen sich auf Version 5.16.0.

Erste Version: 30. August 2018

### Upgrades

- Flink 1.5.2
- HBase 1.4.6
- Presto 0.206

### Neue Features

- Unterstützung für Tensorflow hinzugefügt. Weitere Informationen finden Sie unter [TensorFlow](#).

## Änderungen, Verbesserungen und behobene Probleme

- JupyterHub
  - Unterstützung für Notebook-Persistenz in Amazon S3 hinzugefügt. Weitere Informationen finden Sie unter [Konfigurieren von Persistenz für Notebooks in Amazon S3](#).
- Spark
  - Unterstützung für [S3 Select](#) hinzugefügt. Weitere Informationen finden Sie unter [S3 Select mit Spark zur Verbesserung der Leistung bei Abfragen verwenden](#).
- Die Probleme mit den Cloudwatch-Metriken und des automatischen Skalierungsfeatures in Amazon-EMR-Version 5.14.0, 5.15.0 oder 5.16.0 wurden behoben.

## Bekannte Probleme

- Wenn Sie einen kerberisierten Cluster mit installiertem Livy erstellen, schlägt Livy mit der Fehlermeldung fehl, dass die einfache Authentifizierung nicht aktiviert ist. Ein Neustart des Livy-Servers behebt das Problem. Um das Problem zu umgehen, fügen Sie während der Clustererstellung einen Schritt `sudo restart livy-server` hinzu, der auf dem Primärknoten ausgeführt wird.
- Wenn Sie ein benutzerdefiniertes Amazon-Linux-AMI verwenden, das auf einem Amazon-Linux-AMI mit einem Erstellungsdatum vom 8.11.2018 basiert, kann der Oozie-Server nicht gestartet werden. Wenn Sie Oozie verwenden, erstellen Sie ein benutzerdefiniertes AMI, das auf einer Amazon-Linux-AMI-ID mit einem anderen Erstellungsdatum basiert. Sie können den folgenden AWS CLI Befehl verwenden, um eine Liste der Image-IDs für alle HVM-Amazon-Linux-AMIs mit einer Version 2018.03 zusammen mit dem Veröffentlichungsdatum zurückzugeben, sodass Sie ein geeignetes Amazon Linux-AMI als Basis auswählen können. MyRegion Ersetzen Sie es durch Ihre Regionskennung, z. B. us-west-2.

```
aws ec2 --region MyRegion describe-images --owner amazon --query 'Images[?
Name!=`null`][?starts_with(Name, `amzn-ami-hvm-2018.03`) == `true`].
[CreationDate,ImageId,Name]' --output text | sort -rk1
```

## Version 5.16.0

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.16.0. Änderungen beziehen sich auf Version 5.15.0.

Erste Version: 19. Juli 2018

### Upgrades

- Hadoop 2.8.4
- Flink 1.5.0
- Livy 0.5.0
- MXNet 1.2.0
- Phoenix 4.14.0
- Presto 0.203
- Spark 2.3.1

- AWS SDK for Java 1.11.336
- CUDA 9.2
- Redshift-JDBC-Treiber 1.2.15.1025

## Änderungen, Verbesserungen und behobene Probleme

- HBase
  - [HBASE-20723](#) wurde rückportiert
- Presto
  - Konfigurationsänderungen zur Unterstützung der LDAP-Authentifizierung. Weitere Informationen finden Sie unter [LDAP-Authentifizierung für Presto in Amazon EMR verwenden](#).
- Spark
  - Apache-Spark-Version 2.3.1, verfügbar ab Amazon-EMR-Version 5.16.0, adressiert [CVE-2018-8024](#) und [CVE-2018-1334](#). Wir empfehlen, dass Sie frühere Versionen von Spark zu Spark-Version 2.3.1 oder höher migrieren.

## Bekannte Probleme

- Diese Release-Version unterstützt die Instance-Typen c1.medium oder m1.small nicht. Cluster, die einen dieser Instance-Typen verwenden, können nicht gestartet werden. Um das Problem zu umgehen, geben Sie einen anderen Instance-Typ an oder verwenden Sie eine andere Release-Version.
- Wenn Sie einen kerberisierten Cluster mit installiertem Livy erstellen, schlägt Livy mit der Fehlermeldung fehl, dass die einfache Authentifizierung nicht aktiviert ist. Ein Neustart des Livy-Servers behebt das Problem. Um das Problem zu umgehen, fügen Sie während der Clustererstellung einen Schritt `sudo restart livy-server` hinzu, der auf dem Primärknoten ausgeführt wird.
- Nach dem Neustart des Primärknotens oder des Instance-Controllers werden die CloudWatch Metriken nicht erfasst und die automatische Skalierungsfunktion ist in Amazon EMR Version 5.14.0, 5.15.0 oder 5.16.0 nicht verfügbar. Dieses Problem wurde in Amazon EMR 5.17.0 behoben.

## Version 5.15.0

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.15.0. Änderungen beziehen sich auf Version 5.14.0.

Erste Version: 21. Juni 2018

### Upgrades

- Upgrade auf HBase 1.4.4
- Upgrade auf Hive 2.3.3
- Upgrade auf Hue 4.2.0
- Upgrade auf Oozie 5.0.0
- Upgrade auf ZooKeeper 3.4.12.
- SDK AWS auf 1.11.333 aktualisiert

### Änderungen, Verbesserungen und behobene Probleme

- Hive
  - [HIVE-18069](#) wurde rückportiert
- Hue
  - Hue wurde aktualisiert, um sich korrekt bei Livy zu authentifizieren, wenn Kerberos aktiviert ist. Livy wird jetzt unterstützt, wenn Kerberos mit Amazon EMR verwendet wird.
- JupyterHub
  - Aktualisiert, JupyterHub sodass Amazon EMR standardmäßig LDAP-Clientbibliotheken installiert.
  - Ein Fehler im Skript, das selbstsignierte Zertifikate generiert, wurde behoben.

### Bekannte Probleme

- Diese Release-Version unterstützt die Instance-Typen c1.medium oder m1.small nicht. Cluster, die einen dieser Instance-Typen verwenden, können nicht gestartet werden. Um das Problem zu umgehen, geben Sie einen anderen Instance-Typ an oder verwenden Sie eine andere Release-Version.

- Nach dem Neustart des Primärknotens oder des Instance-Controllers werden die CloudWatch Metriken nicht erfasst und die automatische Skalierungsfunktion ist in Amazon EMR Version 5.14.0, 5.15.0 oder 5.16.0 nicht verfügbar. Dieses Problem wurde in Amazon EMR 5.17.0 behoben.

## Version 5.14.1

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.14.1. Änderungen beziehen sich auf Version 5.14.0.

Veröffentlichungsdatum: 17. Oktober 2018

Das Standard-AMI für Amazon EMR wurde aktualisiert, um potenzielle Sicherheitslücken zu schließen.

## Version 5.14.0

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.14.0. Änderungen beziehen sich auf Version 5.13.0.

Erste Version: 4. Juni 2018

### Upgrades

- Upgrade auf Apache Flink 1.4.2
- Apache MXNet auf 1.1.0 aktualisiert
- Apache Sqoop auf 1.4.7 aktualisiert

### Neue Features

- Unterstützung JupyterHub hinzugefügt. Weitere Informationen finden Sie unter [JupyterHub](#).

### Änderungen, Verbesserungen und behobene Probleme

- EMRFS
  - Die UserAgent-Zeichenfolge in Anfragen an Amazon S3 wurde aktualisiert und enthält nun die Benutzer- und Gruppeninformationen des aufrufenden Prinzipals. Dies kann zusammen mit AWS CloudTrail Protokollen für eine umfassendere Nachverfolgung von Anfragen verwendet werden.

- HBase
  - [HBASE-20447](#) ist enthalten, das ein Problem behebt, das zu Cache-Problemen führen kann, insbesondere bei geteilten Regionen.
- MXnet
  - OpenCV-Plattformversion.
- Spark
  - Wenn Spark Parquet-Dateien mithilfe von EMRFS an einen Amazon S3 S3-Speicherort schreibt, wurde der FileOutputCommitter Algorithmus aktualisiert, sodass er Version 2 statt Version 1 verwendet. Dies reduziert die Anzahl der Umbenennungen, was die Anwendungsleistung verbessert. Diese Änderung beeinträchtigt nicht:
    - Andere Anwendungen als Spark.
    - Anwendungen, die in andere Dateisysteme schreiben, wie HDFS (die immer noch Version 1 von verwenden). FileOutputCommitter
    - Anwendungen, die andere Ausgabeformate wie Text oder CSV verwenden, die bereits EMRFS Direct Write verwenden.

## Bekannte Probleme

- JupyterHub
  - Die Verwendung von Konfigurationsklassifizierungen zur Einrichtung JupyterHub und individuellen Einrichtung von Jupyter-Notebooks beim Erstellen eines Clusters wird nicht unterstützt. Bearbeiten Sie die Dateien `jupyterhub_config.py` und `jupyter_notebook_config.py` für jeden Benutzer manuell. Weitere Informationen finden Sie unter [Konfiguration JupyterHub](#).
  - JupyterHub kann auf Clustern innerhalb eines privaten Subnetzes nicht gestartet werden und schlägt mit der Meldung `fehl. Error: ENOENT: no such file or directory, open '/etc/jupyter/conf/server.crt'` Dies wird durch einen Fehler im Skript verursacht, das selbstsignierte Zertifikate generiert. Verwenden Sie die folgende Problemumgehung zum Generieren selbstsignierter Zertifikate. Alle Befehle werden ausgeführt, während Sie mit dem Primärknoten verbunden sind.
    1. Kopieren Sie das Skript zur Zertifikatsgenerierung aus dem Container auf den Primärknoten:

```
sudo docker cp jupyterhub:/tmp/gen_self_signed_cert.sh ./
```

2. Verwenden Sie einen Texteditor, um Zeile 23 zu ändern, um den öffentlichen Hostnamen in den lokalen Hostnamen zu ändern, wie unten gezeigt:



```
local hostname=$(curl -s $EC2_METADATA_SERVICE_URI/local-hostname)
```

3. Führen Sie das Skript aus, um selbstsignierte Zertifikate zu generieren:

```
sudo bash ./gen_self_signed_cert.sh
```

4. Verschieben Sie die Zertifikatsdateien, die das Skript generiert, in das `/etc/jupyter/conf/`-Verzeichnis:

```
sudo mv /tmp/server.crt /tmp/server.key /etc/jupyter/conf/
```

Sie können `tail` die `jupyter.log` Datei überprüfen, um zu überprüfen, ob sie JupyterHub neu gestartet wurde und einen 200-Antwortcode zurückgibt. Beispielsweise:

```
tail -f /var/log/jupyter/jupyter.log
```

Dadurch sollte eine Antwort ähnlich der folgenden ausgegeben:

```
# [I 2018-06-14 18:56:51.356 JupyterHub app:1581] JupyterHub is now running at
https://:9443/
# 19:01:51.359 - info: [ConfigProxy] 200 GET /api/routes
```

- Nach dem Neustart des Primärknotens oder des Instance-Controllers werden die CloudWatch Metriken nicht erfasst und die automatische Skalierungsfunktion ist in Amazon EMR Version 5.14.0, 5.15.0 oder 5.16.0 nicht verfügbar. Dieses Problem wurde in Amazon EMR 5.17.0 behoben.

## Version 5.13.0

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.13.0. Änderungen beziehen sich auf Version 5.12.0.

### Upgrades

- Upgrade auf Spark 2.3.0
- Upgrade auf HBase 1.4.2
- Upgrade auf Presto 0.194
- Auf 1.11.297 aktualisiert AWS SDK for Java

## Änderungen, Verbesserungen und behobene Probleme

- Hive
  - [HIVE-15436](#) wurde rückportiert. Verbesserte Hive-APIs, sodass nur Ansichten zurückgegeben werden.

## Bekannte Probleme

- MXNet hat derzeit keine OpenCV-Bibliotheken.

## Version 5.12.2

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.12.2. Änderungen beziehen sich auf Version 5.12.1

Erste Version: 29. August 2018

## Änderungen, Verbesserungen und behobene Probleme

- Diese Version behebt eine potenzielle Schwachstellen.

## Version 5.12.1

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.12.1. Änderungen beziehen sich auf Version 5.12.0.

Erste Version: 29. März 2018

## Änderungen, Verbesserungen und behobene Probleme

- Der Amazon-Linux-Kernel des standardmäßigen Amazon-Linux-AMI für Amazon EMR wurde aktualisiert, um potenzielle Schwachstellen zu schließen.

## Version 5.12.0

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.12.0. Änderungen beziehen sich auf Version 5.11.1.

## Upgrades

- AWS SDK for Java 1.11.238 ⇒ 1.11.267. Weitere Informationen finden Sie im [AWS SDK for Java Change Log](#) on GitHub.
- Hadoop 2.7.3 ⇒ 2.8.3. Weitere Informationen finden Sie unter [Apache Hadoop im Releases](#).
- Flink 1.3.2 ⇒ 1.4.0. Weitere Informationen finden Sie in der [Versionsankündigung zu Apache Flink 1.4.0](#).
- HBase 1.3.1 ⇒ 1.4.0. Weitere Informationen finden Sie in der [HBase Versionsankündigung](#)..
- Hue 4.0.1 ⇒ 4.1.0. Weitere Informationen finden Sie in den [Versionshinweisen](#).
- MxNet 0.12.0 ⇒ 1.0.0. Weitere Informationen finden Sie im [MXNet Change Log](#) on GitHub.
- Presto 0.187 ⇒ 0.188. Weitere Informationen finden Sie in den [Versionshinweisen](#).

## Änderungen, Verbesserungen und behobene Probleme

- Hadoop
  - Die `yarn.resourcemanager.decommissioning.timeout`-Eigenschaft wurde zu `yarn.resourcemanager.nodemanager-graceful-decommission-timeout-secs` geändert. Sie können diese Eigenschaft verwenden, um die Cluster-Skalierung anzupassen. Weitere Informationen finden Sie unter [Cluster Sale down](#) im Verwaltungshandbuch für Amazon EMR.
  - Die Hadoop-CLI hat die `-d`-Option zum Befehl `cp` (kopieren) hinzugefügt, der das direkte Kopieren spezifiziert. Sie können dies verwenden, um die Erstellung einer `.COPYING`-Zwischendatei zu vermeiden, wodurch das Kopieren von Daten zwischen Amazon S3 beschleunigt wird. Weitere Informationen finden Sie unter [HADOOP-12384](#).
- Pig
  - Die `pig-env`-Konfigurationsklassifizierung wurde hinzugefügt, die die Konfiguration der Pig-Umgebungseigenschaften vereinfacht. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).
- Presto
  - Die Konfigurationsklassifizierung `presto-connector-redshift` wurde hinzugefügt, mit der Sie Werte in der Presto `redshift.properties` konfigurieren können. Weitere Informationen finden Sie unter [Redshift Connector](#) in der Presto-Dokumentation und [Anwendungen konfigurieren](#).

- Presto-Unterstützung für EMRFS wurde hinzugefügt und ist die Standardkonfiguration. Frühere Amazon EMR-Versionen verwendeten PrestoS3FileSystem, was die einzige Option war. Weitere Informationen finden Sie unter [EMRFS- und PrestoS3-Konfiguration FileSystem](#).

#### Note

Wenn Sie die zugrunde liegenden Daten in Amazon S3 mit Amazon-EMR-Version 5.12.0 abfragen, können Presto-Fehler auftreten. Dies liegt daran, dass Presto die Konfigurationsklassifizierungswerte von `emrfs-site.xml` nicht abrufen. Um dieses Problem zu umgehen, erstellen Sie ein `emrfs`-Unterverzeichnis unter `usr/lib/presto/plugin/hive-hadoop2/` und erstellen einen Symlink `usr/lib/presto/plugin/hive-hadoop2/emrfs` zu der vorhandenen Datei `/usr/share/aws/emr/emrfs/conf/emrfs-site.xml`. Starten Sie dann den Presto-Server-Prozess neu (`sudo presto-server stop` gefolgt von `sudo presto-server start`).

- Spark
  - [SPARK-22036](#) rückportiert: Multiplikation gibt manchmal Null zurück. BigDecimal

#### Bekanntes Probleme

- MXNet enthält keine OpenCV-Bibliotheken.
- SparkR ist nicht für Cluster verfügbar, die mit einem benutzerdefinierten AMI erstellt wurden, da R nicht standardmäßig auf Clusterknoten installiert ist.

## Version 5.11.3

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.11.3. Änderungen beziehen sich auf Version 5.11.2.

Erste Version: 18. Juli 2019

#### Änderungen, Verbesserungen und behobene Probleme

- Das standardmäßige Amazon Linux AMI für Amazon EMR wurde aktualisiert, sodass es wichtige Sicherheitsupdates für den Linux-Kernel enthält, darunter das „TCP SACK Denial of Service Issue“ ([AWS-2019-005](#)).

## Version 5.11.2

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.11.2. Änderungen beziehen sich auf Version 5.11.1.

Erste Version: 29. August 2018

Änderungen, Verbesserungen und behobene Probleme

- Diese Version behebt eine potenzielle Schwachstellen.

## Version 5.11.1

Die folgenden Versionshinweise enthalten Informationen zur Version 5.11.1 von Amazon EMR. Änderungen beziehen sich auf Amazon-EMR-Version 5.11.0.

Erste Version: 22. Januar 2018

Änderungen, Verbesserungen und behobene Probleme

- Der Amazon-Linux-Kernel des standardmäßigen Amazon-Linux-AMI für Amazon EMR wurde aktualisiert, um Schwachstellen im Zusammenhang mit spekulativer Ausführung zu beheben (CVE-2017-5715, CVE-2017-5753 und CVE-2017-5754). Weitere Informationen finden Sie unter <https://aws.amazon.com/security/security-bulletins/AWS-2018-013/>.

## Bekannte Probleme

- MXNet enthält keine OpenCV-Bibliotheken.
- Hive 2.3.2 setzt `hive.compute.query.using.stats=true` standardmäßig. Dies bewirkt, dass Abfragen Daten aus vorhandenen Statistiken und nicht direkt aus den Daten abrufen, da diese unübersichtlich sein können. Wenn Sie beispielsweise über eine Tabelle mit `hive.compute.query.using.stats=true` verfügen und neue Dateien in die Tabelle LOCATION hochladen, gibt die Abfrage `SELECT COUNT(*)` in der Tabelle die Anzahl aus den Statistiken zurück, anstatt die hinzugefügten Zeilen abzurufen.

Um dieses Problem zu umgehen, verwenden Sie den Befehl `ANALYZE TABLE`, um neue Statistiken zu sammeln, oder legen Sie `hive.compute.query.using.stats=false` fest. Weitere Informationen finden Sie unter [Statistics in Hive](#) in der Apache Hive-Dokumentation.

## Version 5.11.0

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.11.0. Änderungen beziehen sich auf Amazon-EMR;-Version 5.10.0.

### Upgrades

Die folgenden Anwendungen und Komponenten wurden in dieser Version aktualisiert, um die folgenden Versionen einzufügen.

- Hive 2.3.2
- Spark 2.2.1
- SDK für Java 1.11.238

### Neue Features

- Spark
  - Die Einstellung `spark.decommissioning.timeout.threshold` wurde hinzugefügt, was das Verhalten von Spark bei der Außerbetriebnahme bei Verwendung von Spot-Instances verbessert. Weitere Informationen finden Sie unter [Verhalten der Außerbetriebnahme von Knoten konfigurieren](#).
  - Die `aws-sagemaker-spark-sdk` Komponente wurde zu Spark hinzugefügt, die Amazon SageMaker Spark und zugehörige Abhängigkeiten für die Spark-Integration mit [Amazon](#) installiert SageMaker. Sie können Amazon SageMaker Spark verwenden, um Spark-Pipelines für maschinelles Lernen (ML) mithilfe von SageMaker Amazon-Stufen zu erstellen. Weitere Informationen finden Sie in der [SageMaker Spark-Readme-Datei](#) GitHub und unter [Using Apache Spark with Amazon SageMaker](#) im Amazon SageMaker Developer Guide.

### Bekanntes Problem

- MXNet enthält keine OpenCV-Bibliotheken.
- Hive 2.3.2 setzt `hive.compute.query.using.stats=true` standardmäßig. Dies bewirkt, dass Abfragen Daten aus vorhandenen Statistiken und nicht direkt aus den Daten abrufen, da diese unübersichtlich sein können. Wenn Sie beispielsweise über eine Tabelle mit `hive.compute.query.using.stats=true` verfügen und neue Dateien in die Tabelle

LOCATION hochladen, gibt die Abfrage `SELECT COUNT(*)` in der Tabelle die Anzahl aus den Statistiken zurück, anstatt die hinzugefügten Zeilen abzurufen.

Um dieses Problem zu umgehen, verwenden Sie den Befehl `ANALYZE TABLE`, um neue Statistiken zu sammeln, oder legen Sie `hive.compute.query.using.stats=false` fest. Weitere Informationen finden Sie unter [Statistics in Hive](#) in der Apache Hive-Dokumentation.

## Version 5.10.0

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.10.0. Änderungen beziehen sich auf Amazon-EMR-Version 5.9.0.

### Upgrades

Die folgenden Anwendungen und Komponenten wurden in dieser Version aktualisiert, um die folgenden Versionen einzufügen.

- AWS SDK for Java 1.11.221
- Hive 2.3.1
- Presto 0.187

### Neue Features

- Die Kerberos-Authentifizierung wird jetzt unterstützt. Weitere Informationen finden Sie unter [Verwendung der Kerberos-Authentifizierung](#) im Verwaltungshandbuch für Amazon EMR.
- Eine Unterstützung für IAM-Rollen für EMRFS-Anfragen an Amazon S3 wurde hinzugefügt. Weitere Informationen finden Sie unter [Konfigurieren von IAM-Rollen für EMRFS Versionen für Amazon S3](#) im Verwaltungshandbuch für Amazon EMR.
- Unterstützung für GPU-basierte P2- und P3-Instance-Typen hinzugefügt. Weitere Informationen finden Sie unter [Amazon-EC2-P2-Instances](#) und [Amazon-EC2-P3-Instances](#). NVIDIA-Treiber 384.81 und CUDA-Treiber 9.0.176 sind auf diesen Instance-Typen standardmäßig installiert.
- Unterstützung für [Apache MXNet](#) hinzugefügt.

### Änderungen, Verbesserungen und behobene Probleme

- Presto

- Unterstützung für die Verwendung des AWS Glue-Datenkatalogs als Standard-Hive-Metastore hinzugefügt. Weitere Informationen finden Sie unter [Presto mit dem AWS Glue-Datenkatalog verwenden](#).
- Unterstützung von [koordinatenbasierten Funktionen](#) hinzugefügt.
- Unterstützung eines [Überlaufs auf die Festplatte](#) für Joins hinzugefügt.
- Unterstützung für den [Redshift-Connector](#) hinzugefügt.
- Spark
  - Rückportierung von [SPARK-20640](#), mit der der rpc-Timeout und die erneuten Versuche für das Shuffling von Registrierungswerten unter Verwendung von `spark.shuffle.registration.timeout` bzw. `spark.shuffle.registration.maxAttempts` konfiguriert werden können.
  - [SPARK-21549](#) wurde rückportiert, wodurch ein Fehler behoben wird, der beim Schreiben von benutzerdefinierten Speicherorten in Nicht-HDFS-Speicherorte auftritt. `OutputFormat`
- Rückportierung von [Hadoop-13270](#)
- Die Bibliotheken Numpy, Scipy und Matplotlib wurden aus der grundlegenden Amazon-EMR-AMI entfernt. Wenn diese Bibliotheken für Ihre Anwendung erforderlich sind, werden sie im folgenden Repository bereitgestellt, sodass Sie sie unter Verwendung einer Bootstrap-Aktion mit `yum install` auf allen Knoten installieren können.
- Die Amazon-EMR-Basis-AMI enthält keine Anwendungs-RPM-Paketen mehr, sodass die RPM-Pakete nicht mehr auf Cluster-Knoten vorhanden sind. Benutzerdefinierte AMIs und die Amazon EMR Basis-AMI von Amazon S3 verweisen jetzt auf das RPM-Paket-Repository in .
- Aufgrund der Einführung von der sekundenweisen Abrechnung in Amazon EC2 ist die Standardeinstellung für Verhalten beim Herunterskalieren jetzt Nach Abschluss der Aufgabe beenden anstelle von Zu einem Zeitpunkt auf der Instance beenden. Weitere Informationen finden Sie unter [Konfigurieren eines Cluster-Scale-Downs](#).

## Bekannte Probleme

- MXNet enthält keine OpenCV-Bibliotheken.
- Hive 2.3.1 setzt `hive.compute.query.using.stats=true` standardmäßig. Dies bewirkt, dass Abfragen Daten aus vorhandenen Statistiken und nicht direkt aus den Daten abrufen, da diese unübersichtlich sein können. Wenn Sie beispielsweise über eine Tabelle mit `hive.compute.query.using.stats=true` verfügen und neue Dateien in die Tabelle



LOCATION hochladen, gibt die Abfrage `SELECT COUNT(*)` in der Tabelle die Anzahl aus den Statistiken zurück, anstatt die hinzugefügten Zeilen abzurufen.

Um dieses Problem zu umgehen, verwenden Sie den Befehl `ANALYZE TABLE`, um neue Statistiken zu sammeln, oder legen Sie `hive.compute.query.using.stats=false` fest. Weitere Informationen finden Sie unter [Statistics in Hive](#) in der Apache Hive-Dokumentation.

## Version 5.9.0

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.9.0. Änderungen beziehen sich auf Amazon-EMR-Version 5.8.0.

Veröffentlichungsdatum: 5. Oktober 2017

Letzte Aktualisierung der Funktionen: 12. Oktober 2017

### Upgrades

Die folgenden Anwendungen und Komponenten wurden in dieser Version aktualisiert, um die folgenden Versionen einzufügen.

- AWS SDK for Java Version 1.11.183
- Flink 1.3.2
- Hue 4.0.1
- Pig 0.17.0
- Presto 0.184

### Neue Features

- Livy-Support hinzugefügt (Version 0.4.0-incubating). Weitere Informationen finden Sie unter [Apache Livy](#).
- Support für Hue-Notebooks für Spark hinzugefügt.
- Support für die i3-Serie für Amazon-EC2-Instances hinzugefügt (12. Oktober 2017).

### Änderungen, Verbesserungen und behobene Probleme

- Spark

- Neue Funktionen hinzugefügt, die sicherstellen, dass Spark die Knotenbeendigung infolge einer manuellen Größenänderung oder einer Anforderung einer Auto Scaling-Richtlinie zuverlässiger ausführt. Weitere Informationen finden Sie unter [Verhalten der Außerbetriebnahme von Knoten konfigurieren](#).
- Anstelle von 3DES wird SSL-Verschlüsselung bei der Übertragung für den Blocktransferservice verwendet, was die Leistung bei der Verwendung von Amazon-EC2-Instance-Typen mit AES-NI verbessert.
- [SPARK-21494](#) rückportiert.
- Zeppelin
  - [ZEPPELIN-2377](#) rückportiert.
- HBase
  - Patch [HBASE-18533 hinzugefügt, der zusätzliche Werte für die HBase-Konfiguration](#) unter Verwendung der Konfigurationsklassifizierung ermöglicht. BucketCache `hbase-site`
- Hue
  - AWS Glue Data Catalog-Unterstützung für den Hive-Abfrageeditor in Hue hinzugefügt.
  - Hauptbenutzer in Hue können standardmäßig alle auf Dateien zugreifen, auf die Amazon-EMR-IAM-Rollen zugreifen dürfen. Neu erstellte Benutzer verfügen nicht automatisch über Zugriffsberechtigungen auf den Amazon-S3-Filebrowser und für ihre Gruppen müssen die `filebrowser.s3_access`-Berechtigungen aktiviert sein.
- Behob einen Fehler, der dazu führte, dass die mit AWS Glue Data Catalog erstellten, zugrunde liegenden JSON-Daten nicht verfügbar waren.

## Bekannte Probleme

- Der Cluster-Start schlägt fehl, wenn alle Anwendungen installiert und die Standardgröße des Amazon-EBS-Stamm-Volumes nicht geändert wurde. Um das Problem zu umgehen, verwenden Sie den `aws emr create-cluster` Befehl von AWS CLI und geben Sie einen größeren `--ebs-root-volume-size` Parameter an.
- Hive 2.3.0 setzt `hive.compute.query.using.stats=true` standardmäßig. Dies bewirkt, dass Abfragen Daten aus vorhandenen Statistiken und nicht direkt aus den Daten abrufen, da diese unübersichtlich sein können. Wenn Sie beispielsweise über eine Tabelle mit `hive.compute.query.using.stats=true` verfügen und neue Dateien in die Tabelle LOCATION hochladen, gibt die Abfrage `SELECT COUNT(*)` in der Tabelle die Anzahl aus den Statistiken zurück, anstatt die hinzugefügten Zeilen abzurufen.

Um dieses Problem zu umgehen, verwenden Sie den Befehl `ANALYZE TABLE`, um neue Statistiken zu sammeln, oder legen Sie `hive.compute.query.using.stats=false` fest. Weitere Informationen finden Sie unter [Statistics in Hive](#) in der Apache Hive-Dokumentation.

## Version 5.8.2

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.8.2. Änderungen beziehen sich auf Version 5.8.1.

Erste Version: 29. März 2018

Änderungen, Verbesserungen und behobene Probleme

- Der Amazon-Linux-Kernel des standardmäßigen Amazon-Linux-AMI für Amazon EMR wurde aktualisiert, um potenzielle Schwachstellen zu schließen.

## Version 5.8.1

Die folgenden Versionshinweise enthalten Informationen zur Version 5.8.1 von Amazon EMR. Änderungen beziehen sich auf Amazon-EMR-Version 5.8.0.

Erste Version: 22. Januar 2018

Änderungen, Verbesserungen und behobene Probleme

- Der Amazon-Linux-Kernel des standardmäßigen Amazon-Linux-AMI für Amazon EMR wurde aktualisiert, um Schwachstellen im Zusammenhang mit spekulativer Ausführung zu beheben (CVE-2017-5715, CVE-2017-5753 und CVE-2017-5754). Weitere Informationen finden Sie unter <https://aws.amazon.com/security/security-bulletins/AWS-2018-013/>.

## Version 5.8.0

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.8.0. Änderungen beziehen sich auf Amazon-EMR-Version 5.7.0.

Erste Version: 10. August 2017

Letzte Aktualisierung der Funktionen: 25. September 2017

## Upgrades

Die folgenden Anwendungen und Komponenten wurden in dieser Version aktualisiert, um die folgenden Versionen einzufügen:

- AWS SDK 1.11.160
- Flink 1.3.1
- Hive 2.3.0. Weitere Informationen finden Sie in den [Versionshinweisen](#) auf der Apache-Hive-Website.
- Spark 2.2.0. Weitere Informationen finden Sie in den [Versionshinweisen](#) auf der Apache-Spark-Website.

## Neue Features

- Support zur Anzeige des Anwendungsverlaufs hinzugefügt (25. September 2017). Weitere Informationen finden Sie unter [Anzeigen des Anwendungsverlaufs](#) im Verwaltungshandbuch für Amazon EMR.

## Änderungen, Verbesserungen und behobene Probleme

- Integration mit dem AWS Glue-Datenkatalog
  - Hive und Spark SQL können jetzt AWS Glue Data Catalog als Hive-Metadatenpeicher verwenden. Weitere Informationen finden Sie unter [Verwenden des AWS Glue-Datenkatalogs als Metastore für Hive](#) und [Verwenden Sie den AWS Glue-Datenkatalog als Metastore für Spark SQL](#).
- Neuen Anwendungsverlauf mit Cluster-Details hinzugefügt, über den Verlaufsdaten für YARN-Anwendungen und zusätzliche Details zu Spark-Anwendungen angezeigt werden können. Weitere Informationen finden Sie unter [Anwendungsverlauf anzeigen](#) im Verwaltungshandbuch für Amazon EMR.
- Oozie
  - [OOZIE-2748](#) wurde rückportiert.
- Hue
  - [HUE-5859](#) wurde rückportiert.
- HBase

- Patch zum Verfügbarmachen der Startzeit des HBase-Master-Servers über Java Management Extensions (JMX) mithilfe von `getMasterInitializedTime` hinzugefügt.
- Patch zur Verbesserung der Cluster-Startzeit hinzugefügt.

## Bekannte Probleme

- Der Cluster-Start schlägt fehl, wenn alle Anwendungen installiert und die Standardgröße des Amazon-EBS-Stamm-Volumes nicht geändert wurde. Um das Problem zu umgehen, verwenden Sie den `aws emr create-cluster` Befehl von AWS CLI und geben Sie einen größeren Parameter an. `--ebs-root-volume-size`
- Hive 2.3.0 setzt `hive.compute.query.using.stats=true` standardmäßig. Dies bewirkt, dass Abfragen Daten aus vorhandenen Statistiken und nicht direkt aus den Daten abrufen, da diese unübersichtlich sein können. Wenn Sie beispielsweise über eine Tabelle mit `hive.compute.query.using.stats=true` verfügen und neue Dateien in die Tabelle LOCATION hochladen, gibt die Abfrage `SELECT COUNT(*)` in der Tabelle die Anzahl aus den Statistiken zurück, anstatt die hinzugefügten Zeilen abzurufen.

Um dieses Problem zu umgehen, verwenden Sie den Befehl `ANALYZE TABLE`, um neue Statistiken zu sammeln, oder legen Sie `hive.compute.query.using.stats=false` fest. Weitere Informationen finden Sie unter [Statistics in Hive](#) in der Apache Hive-Dokumentation.

- Spark – Bei der Verwendung von Spark tritt bei dem appusher-Daemon bei Spark-Aufträgen mit langer Ausführungsdauer nach einigen Stunden oder Tagen möglicherweise ein Speicherleck-Problem bei dem Datei-Handler auf. Um das Problem zu beheben, stellen Sie eine Verbindung mit dem Master-Knoten her und geben `sudo /etc/init.d/appusher stop` ein. Dies beendet den appusher-Daemon, den Amazon EMR automatisch neu startet.
- Application history
  - Es sind keine Verlaufsdaten für beendete Spark-Executors verfügbar.
  - Der Anwendungsverlauf ist nicht für Cluster verfügbar, die eine Sicherheitskonfiguration verwenden, um die Verschlüsselung bei der Übertragung zu aktivieren.

## Version 5.7.0

Die folgenden Versionshinweise enthalten Informationen über Amazon EMR Version 5.7.0. Änderungen beziehen sich auf Amazon-EMR-Version 5.6.0.

Veröffentlichungsdatum: 13. Juli 2017

## Upgrades

- Flink 1.3.0
- Phoenix 4.11.0
- Zeppelin 0.7.2

## Neue Features

- Die Möglichkeit zum Angeben eines benutzerdefinierten Amazon Linux AMI beim Erstellen eines Clusters wurde hinzugefügt. Weitere Informationen finden Sie unter [Verwenden eines benutzerdefinierten AMI](#).

## Änderungen, Verbesserungen und behobene Probleme

- HBase
  - Die Möglichkeit zum Konfigurieren von HBase Read Replica-Cluster wurde hinzugefügt. Weitere Informationen finden Sie unter [Verwenden eines Read-Replica-Clusters](#).
  - Mehrere Fehlerbehebungen und Verbesserungen
- Presto – neue Konfigurationsfunktion `node.properties` hinzugefügt.
- YARN – neue Konfigurationsfunktion `container-log4j.properties` hinzugefügt
- Sqoop – [SQOOP-2880](#) rückportiert und damit einen Parameter eingeführt, mit dem Sie das temporäre Sqoop-Verzeichnis festlegen können.

## Version 5.6.0

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.6.0. Änderungen beziehen sich auf Amazon-EMR-Version 5.5.0.

Veröffentlichungsdatum: 5. Juni 2017

## Upgrades

- Flink 1.2.1
- HBase 1.3.1

- Mahout 0.13.0. Dies ist die erste Version von Mahout, die Spark 2.x in Amazon-EMR-Version 5.0 und höher unterstützt.
- Spark 2.1.1

## Änderungen, Verbesserungen und behobene Probleme

- Presto
  - Die Möglichkeit, SSL-/TLS-geschützte Kommunikation zwischen Presto-Knoten durch Verschlüsselung der Daten während der Übertragung mithilfe einer Sicherheitskonfiguration zu aktivieren, wurde hinzugefügt. Weitere Informationen finden Sie unter [Datenverschlüsselung während der Übertragung](#).
  - [Presto 7661](#) wurde rückportiert, wodurch der Anweisung EXPLAIN ANALYZE die Option VERBOSE hinzugefügt wurde, um detailliertere Low-Level-Statistiken über einen Abfrageplan bereitzustellen.

## Version 5.5.3

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.5.3. Änderungen beziehen sich auf Version 5.5.2.

Erste Version: 29. August 2018

### Änderungen, Verbesserungen und behobene Probleme

- Diese Version behebt eine potenzielle Schwachstellen.

## Version 5.5.2

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.5.2. Änderungen beziehen sich auf Version 5.5.1.

Erste Version: 29. März 2018

### Änderungen, Verbesserungen und behobene Probleme

- Der Amazon-Linux-Kernel des standardmäßigen Amazon-Linux-AMI für Amazon EMR wurde aktualisiert, um potenzielle Schwachstellen zu schließen.

## Version 5.5.1

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.5.1. Änderungen beziehen sich auf Amazon-EMR-Version 5.5.0.

Erste Version: 22. Januar 2018

### Änderungen, Verbesserungen und behobene Probleme

- Der Amazon-Linux-Kernel des standardmäßigen Amazon-Linux-AMI für Amazon EMR wurde aktualisiert, um Schwachstellen im Zusammenhang mit spekulativer Ausführung zu beheben (CVE-2017-5715, CVE-2017-5753 und CVE-2017-5754). Weitere Informationen finden Sie unter <https://aws.amazon.com/security/security-bulletins/AWS-2018-013/>.

## Version 5.5.0

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.5.0. Änderungen beziehen sich auf Amazon-EMR-Version 5.4.0.

Veröffentlichungsdatum: 26. April 2017

### Upgrades

- Hue 3.12
- Presto 0.170
- Zeppelin 0.7.1
- ZooKeeper 3.4.10

### Änderungen, Verbesserungen und behobene Probleme

- Spark
  - Spark-Patch ([SPARK-20115](#)) DAGScheduler wurde korrigiert, um alle verloren gegangenen Shuffle-Blöcke erneut zu berechnen, wenn der Shuffle-Service nicht verfügbar ist wurde auf Version 2.1.0 von Spark rückportiert, die in dieser Version enthalten ist.
- Flink
  - Flink ist nun in Scala 2.11 integriert. Wenn Sie die Scala-API und -Bibliotheken verwenden, empfehlen wir Scala 2.11 für Ihre Projekte.



- Ein Problem wurde behoben, bei dem die Standardeinstellungen `HADOOP_CONF_DIR` und `YARN_CONF_DIR` nicht ordnungsgemäß festgelegt wurden, sodass `start-scala-shell.sh` nicht funktionierte. Außerdem wurde die Möglichkeit hinzugefügt, diese Werte mit `env.hadoop.conf.dir` und `env.yarn.conf.dir` in `/etc/flink/conf/flink-conf.yaml` oder der Konfigurationsklassifizierung `flink-conf` festzulegen.
- Es wurde ein neuer EMR-spezifischer Befehl, `flink-scala-shell` als Wrapper für `start-scala-shell.sh` eingeführt. Wir empfehlen, diese Befehl statt `start-scala-shell` zu verwenden. Der neue Befehl vereinfacht die Ausführung. Beispielsweise startet `flink-scala-shell -n 2` eine Flink Scala-Shell mit einer Aufgabenparallelität von 2.
- Es wurde ein neuer EMR-spezifischer Befehl, `flink-yarn-session` als Wrapper für `yarn-session.sh` eingeführt. Wir empfehlen, diese Befehl statt `yarn-session` zu verwenden. Der neue Befehl vereinfacht die Ausführung. Beispielsweise startet `flink-yarn-session -d -n 2` eine langlebige Flink-Sitzung in einem getrennten Status mit zwei Aufgabenmanagern.
- Das Problem ([FLINK-6125\) Commons httpClient ist in Flink 1.2 nicht mehr schattiert](#)) wurde behoben.
- Presto
  - Die LDAP-Authentifizierung wird jetzt unterstützt. Zur Verwendung von LDAP mit Presto in Amazon EMR; müssen Sie den HTTPS-Zugriff für den Presto-Koordinator (`http-server.https.enabled=true` in `config.properties`) aktivieren. Konfigurationsdetails finden Sie unter [LDAP-Authentifizierung](#) in der Presto-Dokumentation.
  - Unterstützung für `SHOW GRANTS` hinzugefügt.
- Amazon-EMR-Basis-Linux-AMI
  - Amazon-EMR-Versionen basieren jetzt auf Amazon Linux 2017.03. Weitere Informationen finden Sie in den [Amazon-Linux-AMI-2017.03-Versionshinweisen](#).
  - Python 2.6 wurde aus dem Amazon-EMR-Basis-Linux-Image entfernt. Python 2.7 und 3.4 sind standardmäßig installiert. Sie können Python 2.6 bei Bedarf manuell installieren.

## Version 5.4.0

Die folgenden Versionshinweise enthalten Informationen über Amazon-EMR-Version 5.4.0. Änderungen beziehen sich auf Amazon EMR Version 5.3.0.

Veröffentlichungsdatum: 8. März 2017

## Upgrades

Die folgenden Upgrades sind in dieser Version verfügbar:

- Upgrade auf Flink 1.2.0
- Upgrade auf Hbase 1.3.0
- Upgrade auf Phoenix 4.9.0

### Note

Wenn Sie von einer früheren Version von Amazon EMR auf Amazon EMR Version 5.4.0 oder höher upgraden und die sekundäre Indizierung verwenden, aktualisieren Sie lokale Indizes wie in der [Apache-Phoenix-Dokumentation](#) beschrieben. Amazon EMR entfernt die erforderlichen Konfigurationen aus der `hbase-site`-Klassifizierung, die Indizes müssen jedoch neu gefüllt werden. Online- und Offline-Upgrades von Indizes werden unterstützt. Online-Upgrades sind die Standardeinstellung. Dies bedeutet, dass Indizes neu gefüllt werden, während die Initialisierung von Phoenix-Clients Version 4.8.0 oder höher ausgeführt wird. Um Offline-Upgrades festzulegen, geben Sie für die `phoenix.client.localIndexUpgrade`-Konfiguration in der `phoenix-site`-Klassifizierung "false" und dann SSH für den Master-Knoten an, um `psql [zookeeper] -1` auszuführen.

- Upgrade auf Presto 0.166
- Upgrade auf Zeppelin 0.7.0

## Änderungen und Verbesserungen

Die folgenden Änderungen wurden an Amazon-EMR-Versionen mit der Versionsbezeichnung `emr-5.4.0` vorgenommen:

- Es werden jetzt `r4`-Instances unterstützt. Sehen Sie unter: [Amazon-EC2-Instance-Typen](#)

## Version 5.3.1

Die folgenden Versionshinweise enthalten Informationen über Amazon EMR Version 5.3.1. Änderungen beziehen sich auf Amazon EMR Version 5.3.0.

Veröffentlichungsdatum: 7. Februar 2017

Kleinere Änderungen am Backport von Zeppelin-Patches und an der Aktualisierung des Standard-AMI für Amazon EMR.

## Version 5.3.0

Die folgenden Versionshinweise enthalten Informationen über Amazon EMR Version 5.3.0. Änderungen beziehen sich auf Amazon EMR Version 5.2.1.

Veröffentlichungsdatum: 26. Januar 2017

### Upgrades

Die folgenden Upgrades sind in dieser Version verfügbar:

- Upgrade auf Hive 2.1.1
- Upgrade auf Hue 3.11.0
- Upgrade auf Spark 2.1.0
- Upgrade auf Oozie 4.3.0
- Upgrade auf Flink 1.1.4

### Änderungen und Verbesserungen

Die folgenden Änderungen wurden an Amazon-EMR-Versionen mit der Versionsbezeichnung emr-5.3.0 vorgenommen:

- Ein Patch für Hue wurde hinzugefügt, mit dem Sie die Einstellung `interpreters_shown_on_wheel` verwenden können, um zu konfigurieren, was Interpreter in der Notebook-Auswahl als Erstes anzeigen, unabhängig von ihrer Reihenfolge in der Datei `hue.ini`.
- Die Konfigurationsklassifizierung `hive-parquet-logging` wurde hinzugefügt, mit der Sie Werte in der Hive-Datei `parquet-logging.properties` konfigurieren können.

## Version 5.2.2

Die folgenden Versionshinweise enthalten Informationen über Amazon EMR Version 5.2.2. Änderungen beziehen sich auf Amazon EMR Version 5.2.1.

Veröffentlichungsdatum: 2. Mai 2017

## Bekannte Probleme aus den früheren Versionen, die behoben wurden

- [SPARK-194459](#) wurde rückportiert. Damit wurde ein Problem behoben, bei dem beim Lesen aus einer ORC-Tabelle mit CHAR/VARCHAR-Spalten Fehler auftreten konnten.

## Version 5.2.1

Die folgenden Versionshinweise enthalten Informationen über Amazon EMR Version 5.2.1.

Änderungen beziehen sich auf Amazon EMR Version 5.2.0.

Veröffentlichungsdatum: 29. Dezember 2016

## Upgrades

Die folgenden Upgrades sind in dieser Version verfügbar:

- Upgrade auf Presto 0.157.1. Weitere Informationen finden Sie in den [Presto-Versionshinweisen](#) in der Presto-Dokumentation.
- Upgrade auf ZooKeeper 3.4.9. Weitere Informationen finden Sie in den [ZooKeeper Versionshinweisen](#) in der ZooKeeper Apache-Dokumentation.

## Änderungen und Verbesserungen

Die folgenden Änderungen wurden an Amazon-EMR-Versionen mit der Versionsbezeichnung emr-5.2.1 vorgenommen:

- Unterstützung hinzugefügt für den Amazon-EC2-Instance-Typ m4.16xlarge in Amazon-EMR-Version 4.8.3 und höher, ausgenommen 5.0.0, 5.0.3 und 5.2.0.
- Amazon-EMR-Versionen basieren jetzt auf Amazon Linux 2016.09. Weitere Informationen finden Sie unter <https://aws.amazon.com/amazon-linux-ami/2016.09-release-notes/>.
- Der Speicherort von Flink- und YARN-Konfigurationspfaden wird nun standardmäßig in `/etc/default/flink` festgelegt. Sie brauchen die Umgebungsvariablen `FLINK_CONF_DIR` und `HADOOP_CONF_DIR` bei der Ausführung des Treiberskripts `flink` oder `yarn-session.sh` zum Starten von Flink-Aufträgen nicht festlegen.
- Unterstützung für FlinkKinesisConsumer Klassen hinzugefügt.

## Bekannte Probleme aus den früheren Versionen, die behoben wurden

- Es wurde ein Problem in Hadoop behoben, bei dem der ReplicationMonitor Thread aufgrund eines Wettlaufs zwischen Replikation und Löschung derselben Datei in einem großen Cluster für lange Zeit hängen bleiben konnte.
- Es wurde ein Problem behoben, bei dem ControlledJob #toString mit einer Nullzeiger-Ausnahme (NPE) fehlschlug, wenn der Jobstatus nicht erfolgreich aktualisiert wurde.

## Version 5.2.0

Die folgenden Versionshinweise enthalten Informationen über Amazon EMR Version 5.2.0. Änderungen beziehen sich auf Amazon EMR Version 5.1.0.

Veröffentlichungsdatum: 21. November 2016

### Änderungen und Verbesserungen

Die folgenden Änderungen und Verbesserungen sind in dieser Version verfügbar:

- Der Amazon-S3-Speichermodus für HBase wurde hinzugefügt.
- Jetzt können Sie einen Amazon-S3-Speicherort für das HBase-Stammverzeichnis angeben. Weitere Informationen finden Sie unter [HBase in Amazon S3](#).

### Upgrades

Die folgenden Upgrades sind in dieser Version verfügbar:

- Upgrade auf Spark 2.0.2

## Bekannte Probleme aus den früheren Versionen, die behoben wurden

- Es wurde ein Problem mit der Beschränkung von /mnt auf 2 TB auf ausschließlichen EBS-Instanztypen behoben.
- Es wurde ein Problem mit Instance-Controller und Logpusher-Protokollen behoben, die an die entsprechenden OUT-Dateien anstelle ihrer üblichen log4j-konfigurierten LOG-Dateien, die stündlich rotieren, ausgegeben wurden. Da die OUT-Dateien nicht rotieren, ist letztendlich die /emr-Partition belegt. Dieses Problem wirkt sich nur auf Hardware Virtual Machine (HVM)-Instanztypen aus.

## Version 5.1.0

Die folgenden Versionshinweise enthalten Informationen über Amazon EMR Version 5.1.0. Änderungen beziehen sich auf Amazon EMR Version 5.0.0.

Veröffentlichungsdatum: 03. November 2016

### Änderungen und Verbesserungen

Die folgenden Änderungen und Verbesserungen sind in dieser Version verfügbar:

- Flink 1.1.3 wird jetzt unterstützt.
- Presto wurde als Option im Notebookabschnitt von Hue hinzugefügt.

### Upgrades

Die folgenden Upgrades sind in dieser Version verfügbar:

- Upgrade auf Hbase 1.2.3
- Upgrade auf Zeppelin 0.6.2

### Bekannte Probleme aus den früheren Versionen, die behoben wurden

- Ein Problem mit Tez-Abfragen in Amazon S3, bei dem ORC-Dateien nicht so gute Leistung zeigten, wie in früheren Amazon-EMR-4.x-Versionen, wurde behoben.

## Version 5.0.3

Die folgenden Versionshinweise enthalten Informationen über Amazon EMR Version 5.0.3. Änderungen beziehen sich auf Amazon EMR Version 5.0.0.

Veröffentlichungsdatum: 24. Oktober 2016

### Upgrades

Die folgenden Upgrades sind in dieser Version verfügbar:

- Upgrade auf Hadoop 2.7.3

- Es wurde ein Upgrade auf Presto 0.152.3 durchgeführt, das Support für die Presto-Webschnittstelle umfasst. Sie können auf die Presto-Webschnittstelle des Presto-Koordinators über Port 8889 zugreifen. Weitere Informationen über die Presto-Webschnittstelle finden Sie unter [Webschnittstelle](#) in der Presto-Dokumentation.
- Upgrade auf Spark 2.0.1
- Amazon-EMR-Versionen basieren jetzt auf Amazon Linux 2016.09. Weitere Informationen finden Sie unter <https://aws.amazon.com/amazon-linux-ami/2016.09-release-notes/>.

## Version 5.0.0

Veröffentlichungsdatum: 27. Juli 2016

### Upgrades

Die folgenden Upgrades sind in dieser Version verfügbar:

- Upgrade auf Hive 2.1
- Upgrade auf Presto 0.150
- Upgrade auf Spark 2.0
- Upgrade auf Hue 3.10.0
- Upgrade auf Pig 0.16.0
- Upgrade auf Tez 0.8.4
- Upgrade auf Zeppelin 0.6.1

### Änderungen und Verbesserungen

Die folgenden Änderungen wurden an Amazon-EMR-Versionen mit der Versionsbezeichnung emr-5.0.0 oder höher vorgenommen:

- Amazon EMR unterstützt die neuesten Open-Source-Versionen von Hive (Version 2.1) und Pig (Version 0.16.0). Wenn Sie Hive oder Pig in der Vergangenheit in Amazon EMR verwendet haben, kann sich dies auf einige Anwendungsfälle auswirken. Weitere Informationen finden Sie unter [Hive](#) und [Pig](#).
- Die Standard-Ausführungs-Engine für Hive und Pig ist jetzt Tez. Um dies zu ändern, bearbeiten Sie die entsprechenden Werte in den Konfigurationsklassifizierungen `hive-site` bzw. `pig-properties`.

- Eine verbesserte Schritt-Debugging-Funktion wurde hinzugefügt. Mit dieser Funktion können Sie die Ursache von Schrittfehlern sehen, wenn der Service die Ursache ermitteln kann. Weitere Informationen finden Sie unter [Verbessertes Schritt-Debuggen](#) im Verwaltungshandbuch für Amazon EMR.
- Anwendungen, die zuvor mit "-Sandbox" endeten, tragen diesen Suffix nicht mehr. Dies kann Ihre Automatisierung, zum Beispiel, wenn Sie Skripts zum Starten von Clustern mit diesen Anwendungen verwenden, unterbrechen. In der folgenden Tabelle sind die Namen von Anwendungen in Amazon EMR 4.7.2 im Vergleich zu Amazon EMR 5.0.0 aufgeführt.

#### Änderungen der Anwendungsnamen

Amazon EMR 4.7.2	Amazon EMR 5.0.0
Oozie-Sandbox	Oozie
Presto-Sandbox	Presto
Sqoop-Sandbox	Sqoop
Zeppelin-Sandbox	Zeppelin
ZooKeeper-Sandbox	ZooKeeper

- Spark ist jetzt für Scala 2.11 kompiliert.
- Java 8 ist jetzt die Standard-JVM. Alle Anwendungen werden mit der Java 8-Laufzeit ausgeführt. Es sind keine Änderungen am Byte-Codeziel der Anwendungen vorgenommen worden. Die meisten Anwendungen zielen weiterhin auf Java 7 ab.
- Zeppelin enthält jetzt Authentifizierungsfunktionen. Weitere Informationen finden Sie unter [Zeppelin](#).
- Es werden jetzt Sicherheitskonfigurationen unterstützt, mit deren Hilfe Sie die Verschlüsselungsoptionen einfacher erstellen und anwenden können. Weitere Informationen finden Sie unter [Datenverschlüsselung](#).

## Version 4.9.5

Die folgenden Versionshinweise enthalten Informationen über Amazon EMR Version 4.9.5. Änderungen beziehen sich auf Version 4.9.4.

Erste Version: 29. August 2018



## Änderungen, Verbesserungen und behobene Probleme

- HBase
  - Diese Version behebt eine potenzielle Schwachstellen.

## Version 4.9.4

Die folgenden Versionshinweise enthalten Informationen über Amazon EMR Version 4.9.4. Änderungen beziehen sich auf Version 4.9.3.

Erste Version: 29. März 2018

## Änderungen, Verbesserungen und behobene Probleme

- Der Amazon-Linux-Kernel des standardmäßigen Amazon-Linux-AMI für Amazon EMR wurde aktualisiert, um potenzielle Schwachstellen zu schließen.

## Version 4.9.3

Die folgenden Versionshinweise enthalten Informationen über Amazon EMR Version 4.9.3. Änderungen beziehen sich auf Amazon EMR Version 4.9.2.

Erste Version: 22. Januar 2018

## Änderungen, Verbesserungen und behobene Probleme

- Der Amazon-Linux-Kernel des standardmäßigen Amazon-Linux-AMI für Amazon EMR wurde aktualisiert, um Schwachstellen im Zusammenhang mit spekulativer Ausführung zu beheben (CVE-2017-5715, CVE-2017-5753 und CVE-2017-5754). Weitere Informationen finden Sie unter <https://aws.amazon.com/security/security-bulletins/AWS-2018-013/>.

## Version 4.9.2

Die folgenden Versionshinweise enthalten Informationen über Amazon EMR Version 4.9.2. Änderungen beziehen sich auf Amazon EMR Version 4.9.1.

Veröffentlichungsdatum: 13. Juli 2017

Kleinere Änderungen, Fehlerbehebungen und Verbesserungen wurden in dieser Version vorgenommen.

## Version 4.9.1

Die folgenden Versionshinweise enthalten Informationen über Amazon EMR Version 4.9.1. Änderungen beziehen sich auf Amazon EMR Version 4.8.4.

Veröffentlichungsdatum: 10. April 2017

Bekannte Probleme aus den früheren Versionen, die behoben wurden

- Rückportierungen von [HIVE-9976](#) und [HIVE-10106](#)
- Es wurde ein Problem in YARN behoben, bei dem eine große Anzahl von Knoten (mehr als 2.000) und Container (mehr als 5.000) einen Fehler wegen zu geringem Speicherplatz verursachte, z. B.: "Exception in thread 'main' java.lang.OutOfMemoryError".

## Änderungen und Verbesserungen

Die folgenden Änderungen wurden an Amazon-EMR-Versionen mit der Versionsbezeichnung emr-4.9.1 vorgenommen:

- Amazon-EMR-Versionen basieren jetzt auf Amazon Linux 2017.03. Weitere Informationen finden Sie unter <https://aws.amazon.com/amazon-linux-ami/2017.03-release-notes/>.
- Python 2.6 wurde aus dem Amazon-EMR-Basis-Linux-Image entfernt. Sie können Python 2.6 bei Bedarf manuell installieren.

## Version 4.8.4

Die folgenden Versionshinweise enthalten Informationen über Amazon EMR Version 4.8.4. Änderungen beziehen sich auf Amazon EMR Version 4.8.3.

Veröffentlichungsdatum: 7. Februar 2017

Kleinere Änderungen, Fehlerbehebungen und Verbesserungen wurden in dieser Version vorgenommen.

## Version 4.8.3

Die folgenden Versionshinweise enthalten Informationen über Amazon EMR Version 4.8.3. Änderungen beziehen sich auf Amazon EMR Version 4.8.2.

Veröffentlichungsdatum: 29. Dezember 2016

### Upgrades

Die folgenden Upgrades sind in dieser Version verfügbar:

- Upgrade auf Presto 0.157.1. Weitere Informationen finden Sie in den [Presto-Versionshinweisen](#) in der Presto-Dokumentation.
- Upgrade auf Spark 1.6.3. Weitere Informationen finden Sie in den [Spark-Versionshinweisen](#) in der Apache-Spark-Dokumentation.
- Auf 3.4.9 aktualisiert. ZooKeeper Weitere Informationen finden Sie in den [ZooKeeper Versionshinweisen](#) in der ZooKeeper Apache-Dokumentation.

### Änderungen und Verbesserungen

Die folgenden Änderungen wurden an Amazon-EMR-Versionen mit der Versionsbezeichnung emr-4.8.3 vorgenommen:

- Unterstützung hinzugefügt für den Amazon-EC2-Instance-Typ m4.16xlarge in Amazon-EMR-Version 4.8.3 und höher, ausgenommen 5.0.0, 5.0.3 und 5.2.0.
- Amazon-EMR-Versionen basieren jetzt auf Amazon Linux 2016.09. Weitere Informationen finden Sie unter <https://aws.amazon.com/amazon-linux-ami/2016.09-release-notes/>.

### Bekannte Probleme aus den früheren Versionen, die behoben wurden

- Es wurde ein Problem in Hadoop behoben, bei dem der ReplicationMonitor Thread aufgrund eines Wettlaufs zwischen Replikation und Löschung derselben Datei in einem großen Cluster für lange Zeit hängen bleiben konnte.
- Es wurde ein Problem behoben, bei dem ControlledJob #toString mit einer Nullzeiger-Ausnahme (NPE) fehlschlug, wenn der Jobstatus nicht erfolgreich aktualisiert wurde.

## Version 4.8.2

Die folgenden Versionshinweise enthalten Informationen über Amazon EMR Version 4.8.2. Änderungen beziehen sich auf Amazon EMR Version 4.8.0.

Veröffentlichungsdatum: 24. Oktober 2016

### Upgrades

Die folgenden Upgrades sind in dieser Version verfügbar:

- Upgrade auf Hadoop 2.7.3
- Es wurde ein Upgrade auf Presto 0.152.3 durchgeführt, das Support für die Presto-Webschnittstelle umfasst. Sie können auf die Presto-Webschnittstelle des Presto-Koordinators über Port 8889 zugreifen. Weitere Informationen über die Presto-Webschnittstelle finden Sie unter [Webschnittstelle](#) in der Presto-Dokumentation.
- Amazon-EMR-Versionen basieren jetzt auf Amazon Linux 2016.09. Weitere Informationen finden Sie unter <https://aws.amazon.com/amazon-linux-ami/2016.09-release-notes/>.

## Version 4.8.0

Veröffentlichungsdatum: 7. September 2016

### Upgrades

Die folgenden Upgrades sind in dieser Version verfügbar:

- Upgrade auf Hbase 1.2.2
- Upgrade auf Presto-Sandbox 0.151
- Upgrade auf Tez 0.8.4
- Upgrade auf Zeppelin-Sandbox 0.6.1

### Änderungen und Verbesserungen

Die folgenden Änderungen wurden an Amazon-EMR-Versionen mit der Versionsbezeichnung emr-4.8.0 vorgenommen:

- Es wurde ein Problem in YARN behoben, ApplicationMaster bei dem versucht wurde, Container zu bereinigen, die nicht mehr existieren, weil ihre Instanzen beendet wurden.
- Die hive-server2-URL für Hive2-Aktionen in den Oozie-Beispielen wurde korrigiert.
- Zusätzliche Presto-Kataloge werden jetzt unterstützt.
- Patches wurden rückportiert: [HIVE-8948](#), [HIVE-12679](#), [HIVE-13405](#), [PHOENIX-3116](#), [HADOOP-12689](#)
- Es werden jetzt Sicherheitskonfigurationen unterstützt, mit deren Hilfe Sie die Verschlüsselungsoptionen einfacher erstellen und anwenden können. Weitere Informationen finden Sie unter [Datenverschlüsselung](#).

## Version 4.7.2

Die folgenden Versionshinweise enthalten Informationen über Amazon EMR 4.7.2.

Veröffentlichungsdatum: 15. Juli 2016

### Features

Die folgenden Funktionen sind in dieser Version verfügbar:

- Upgrade auf Mahout 0.12.2
- Upgrade auf Presto 0.148
- Upgrade auf Spark 1.6.2
- Sie können jetzt eine AWSCredentialsProvider für die Verwendung mit EMRFS erstellen, indem Sie einen URI als Parameter verwenden. Weitere Informationen finden Sie unter [Erstellen eines AWSCredentialsProvider für EMRFS](#).
- EMRFS ermöglicht Benutzern nun die Konfiguration eines benutzerdefinierten DynamoDB -Endpunkts für die Metadaten ihrer konsistenten Ansicht mit der Eigenschaft `fs.s3.consistent.dynamodb.endpoint` in `emrfs-site.xml`.
- Es wurde ein Skript in `/usr/bin` mit dem Namen `spark-example` hinzugefügt, das `/usr/lib/spark/spark/bin/run-example` umschließt, sodass Sie Beispiele direkt ausführen können. Um beispielsweise das SparkPi Beispiel auszuführen, das in der Spark-Distribution enthalten ist, können Sie es `spark-example SparkPi 100` von der Befehlszeile aus ausführen oder es `command-runner.jar` als Schritt in der API verwenden.

## Bekannte Probleme aus den früheren Versionen, die behoben wurden

- Es wurde ein Problem behoben, bei dem Oozie die Datei `spark-assembly.jar` nicht am richtigen Speicherort ablegte, wenn Spark auch installiert war. Dies führte dazu, dass Spark-Anwendungen nicht mit Oozie gestartet werden konnten.
- Es wurde ein Problem mit Spark Log4j-basierter Protokollierung in YARN-Containern behoben.

## Version 4.7.1

Veröffentlichungsdatum: 10. Juni 2016

## Bekannte Probleme aus den früheren Versionen, die behoben wurden

- Es wurde ein Problem behoben, das die Startup-Zeit von Clustern verlängerte, die in einer VPC mit privaten Subnetzen gestartet wurden. Der Fehler wirkte sich nur auf Cluster aus, die mit Amazon EMR Version 4.7.0 gestartet wurden.
- Es wurde ein Problem behoben, bei dem das Auflisten von Dateien in Amazon EMR für Cluster, die mit Amazon EMR Version 4.7.0 gestartet wurden, nicht ordnungsgemäß durchgeführt wurde.

## Version 4.7.0

### Important

Amazon EMR 4.7.0 ist veraltet. Verwenden Sie stattdessen Amazon EMR 4.7.1 oder höher.

Veröffentlichungsdatum: 2. Juni 2016

## Features

Die folgenden Funktionen sind in dieser Version verfügbar:

- Apache Phoenix 4.7.0 wurde hinzugefügt.
- Apache Tez 0.8.3 wurde hinzugefügt.
- Upgrade auf Hbase 1.2.1
- Upgrade auf Mahout 0.12.0

- Upgrade auf Presto 0.147
- Auf AWS SDK for Java 1.10.75 aktualisiert
- Das letzte Flag wurde von der Eigenschaft `mapreduce.cluster.local.dir` in `mapred-site.xml` entfernt, damit Benutzer Pig im lokalen Modus ausführen können.

## Amazon-Redshift-JDBC-Treiber verfügbar auf dem Cluster

Amazon Redshift JDBC-Treiber sind jetzt enthalten unter `/usr/share/aws/redshift/jdbc/`, `/usr/share/aws/redshift/jdbc/RedshiftJDBC41.jar` ist der JDBC 4.1-kompatible Amazon Redshift-Treiber und `/usr/share/aws/redshift/jdbc/RedshiftJDBC4.jar` ist der JDBC-4.0-kompatible Amazon Redshift-Treiber. Weitere Informationen finden Sie unter [Konfigurieren von JDBC-Verbindungen](#) im Amazon-Redshift-Verwaltungshandbuch.

## Java 8

Außer für Presto ist OpenJDK 1.7 das Standard-JDK, das für alle Anwendungen verwendet wird. Es ist jedoch sowohl OpenJDK 1.7 als auch 1.8 installiert. Informationen zum Festlegen von `JAVA_HOME` für Anwendungen finden Sie unter [Konfigurieren von Anwendungen zur Verwendung von Java 8](#).

## Bekannte Probleme aus den früheren Versionen, die behoben wurden

- Es wurde ein Kernel-Problem behoben, das sich signifikant auf die Leistung für durchsatzoptimierte HDD (st1)-EBS-Volumes für Amazon EMR in emr-4.6.0 auswirkte.
- Es wurde ein Problem behoben, bei dem ein Cluster fehlschlug, wenn eine HDFS-Verschlüsselungszone angegeben wurde, ohne Hadoop als Anwendung auszuwählen.
- Die HDFS-Standardschreibrichtlinie wurde von `RoundRobin` in `AvailableSpaceVolumeChoosingPolicy` geändert. Einige Volumes wurden mit der `RoundRobin` Konfiguration nicht richtig genutzt, was zu fehlerhaften Kernknoten und einem unzuverlässigen HDFS führte.
- Es wurde ein Problem mit der EMRFS-Befehlszeilenschnittstelle behoben, das zu einer Ausnahme beim Erstellen der standardmäßigen DynamoDB-Metadatentabelle für konsistente Ansichten führte.
- Es wird ein Deadlock-Problem in EMRFS behoben, das bei mehrteiligen Umbenennungs- und Kopieroperationen auftreten konnte.
- Es wurde ein Problem mit EMRFS behoben, das dazu führte, dass die `CopyPart` Standardgröße 5 MB betrug. Der Standard ist jetzt ordnungsgemäß auf 128 MB eingestellt.

- Es wurde ein Problem mit der Zeppelin-Startkonfiguration behoben, das möglicherweise das Anhalten des Service verhinderte.
- Es wurde ein Problem mit Spark und Zeppelin behoben, das Sie daran hindert, das `s3a://`-URI-Schema zu verwenden, da `/usr/lib/hadoop/hadoop-aws.jar` nicht ordnungsgemäß in den jeweiligen Klassenpfad geladen wurde.
- [HUE-2484](#) wurde rückportiert.
- Es wurde ein [Commit](#) von Hue 3.9.0 (JIRA nicht vorhanden) rückportiert, um ein Problem mit dem HBase-Browserbeispiel zu beheben.
- [HIVE-9073](#) wurde rückportiert.

## Version 4.6.0

Veröffentlichungsdatum: 21. April 2016

### Features

Die folgenden Funktionen sind in dieser Version verfügbar:

- HBase 1.2.0 wurde hinzugefügt.
- ZooKeeper-Sandbox 3.4.8 wurde hinzugefügt.
- Upgrade auf Presto-Sandbox 0.143
- Amazon-EMR-Versionen basieren jetzt auf Amazon Linux 2016.03.0. Weitere Informationen finden Sie unter <https://aws.amazon.com/amazon-linux-ami/2016.03-release-notes/>.

### Problem mit den durchsatzoptimierten HDD (st1)-EBS-Volumen-Typen

Ein Problem im Linux-Kernel Versionen 4.2 und höher wirkte sich signifikant auf die Leistung der durchsatzoptimierten HDD (st1)-EBS-Volumen für EMR aus. Diese Version (emr-4.6.0) verwendet die Kernel-Version 4.4.5 und ist daher betroffen. Es wird empfohlen, emr-4.6.0 nicht zu verwenden, wenn Sie st1-EBS-Volumen einsetzen möchten. Sie können Amazon-EMR-Versionen emr-4.5.0 oder früher mit st1 ohne Auswirkungen verwenden. Außerdem stellen wir die Problembehebung mit späteren Versionen zur Verfügung.

### Python-Standardinstellungen

Python 3.4 ist jetzt standardmäßig installiert, Python 2.7 bleibt jedoch der Systemstandard. Sie können Python 3.4 mit einer Bootstrap-Aktion als Systemstandard konfigurieren. Sie können die



Konfigurations-API verwenden, um den PYSPARK\_PYTHON-Export `/usr/bin/python3.4` in der `spark-env` Klassifizierung auf festzulegen, um die Python-Version zu beeinflussen, die verwendet wird. PySpark

## Java 8

Außer für Presto ist OpenJDK 1.7 das Standard-JDK, das für alle Anwendungen verwendet wird. Es ist jedoch sowohl OpenJDK 1.7 als auch 1.8 installiert. Informationen zum Festlegen von `JAVA_HOME` für Anwendungen finden Sie unter [Konfigurieren von Anwendungen zur Verwendung von Java 8](#).

## Bekannte Probleme aus den früheren Versionen, die behoben wurden

- Es wurde ein Problem behoben, bei dem die Anwendungsbereitstellung manchmal wegen eines generierten Passwort nach dem Zufallsprinzip ausfiel.
- Bisher war `mysql` auf allen Knoten installiert. Jetzt ist es nur auf der Master-Instance installiert und nur dann, wenn die gewählte Anwendung `mysql-server` als Komponente enthält. Derzeit enthalten die folgenden Anwendungen die Komponente `mysql-server`: HCatalog, Hive, Hue, Presto-Sandbox und Sqoop-Sandbox.
- Der Wert `yarn.scheduler.maximum-allocation-vcores` wurde von der Standardeinstellung 32 auf 80 geändert. Dadurch wurde in `emr-4.4.0` eingeführtes Problem behoben, das hauptsächlich mit Spark auftritt bei der Verwendung der Option `maximizeResourceAllocation` in einem Cluster, dessen Core-Instance-Typ zu den wenigen großen Instance-Typen gehört, deren YARN `vcores` höher als 32 festgelegt sind. Und zwar waren `c4.8xlarge`, `cc2.8xlarge`, `hs1.8xlarge`, `i2.8xlarge`, `m2.4xlarge`, `r3.8xlarge`, `d2.8xlarge` oder `m4.10xlarge` von dem Problem betroffen.
- `s3-dist-cp` verwendet jetzt für alle Amazon-S3-Nominierungen EMRFS und keine Stufen eines temporären HDFS-Verzeichnisses mehr.
- Es wurde ein Problem mit der Ausnahmebehandlung für die clientseitigen Verschlüsselung mehrteiliger Uploads behoben.
- Es wurde eine Option hinzugefügt, mit der Benutzer die Amazon-S3-Speicherklasse ändern können. Standardmäßig ist diese Einstellung `STANDARD`. Die Konfigurationsklassifizierungseinstellung `emrfs-site` ist `fs.s3.storageClass` und die möglichen Werte sind `STANDARD`, `STANDARD_IA` und `REDUCED_REDUNDANCY`. Weitere Informationen zu Speicherklassen finden Sie unter [Speicherklassen](#) im Benutzerhandbuch zu Amazon Simple Storage Service.

## Version 4.5.0

Veröffentlichungsdatum: 4. April 2016

### Features

Die folgenden Funktionen sind in dieser Version verfügbar:

- Upgrade auf Spark 1.6.1
- Upgrade auf Hadoop 2.7.2
- Upgrade auf Presto 0.140
- AWS KMS Unterstützung für serverseitige Amazon S3 S3-Verschlüsselung hinzugefügt.

### Bekanntes Probleme aus den früheren Versionen, die behoben wurden

- Es wurde ein Problem behoben, bei dem MySQL- und Apache-Server nicht gestartet wurden, nachdem ein Knoten neu gestartet wurde.
- Es wurde ein Problem behoben, bei dem IMPORT nicht ordnungsgemäß mit nicht partitionierte Tabellen in Amazon S3 funktionierte.
- Es wurde ein Problem mit Presto behoben, bei dem das Staging-Verzeichnis beim Schreiben in Hive-Tabellen /mnt/tmp anstatt /tmp lauten muss.

## Version 4.4.0

Veröffentlichungsdatum: 14. März 2016

### Features

Die folgenden Funktionen sind in dieser Version verfügbar:

- HCatalog 1.0.0 wurde hinzugefügt.
- Sqoop-Sandbox 1.4.6 wurde hinzugefügt.
- Upgrade auf Presto 0.136
- Upgrade auf Zeppelin 0.5.6
- Upgrade auf Mahout 0.11.1
- `dynamicResourceAllocation` wurde standardmäßig aktiviert.

- Es wurde eine Tabelle mit allen Konfigurationsklassifizierungen für die Version hinzugefügt. Weitere Informationen finden Sie in der Konfigurationsklassifizierungstabelle unter [Konfigurieren von Anwendungen](#).

## Bekannte Probleme aus den früheren Versionen, die behoben wurden

- Es wurde ein Problem behoben, bei dem die `maximizeResourceAllocation` Einstellung nicht genug Speicher für ApplicationMaster YARN-Daemons reservierte.
- Es wurde ein Problem mit einer benutzerdefinierten DNS behoben. Wenn Einträge in `resolve.conf` den angegebenen, benutzerdefinierten Einträge vorangestellt werden, können die benutzerdefinierten Einträge nicht aufgelöst werden. Dieses Verhalten wirkte sich auf Cluster in einer VPC aus, wo der Standard-VPC-Nameserver als oberster Eintrag in `resolve.conf` eingefügt wird.
- Es wurde ein Problem behoben, bei dem die Python-Standardversion auf Version 2.7 geändert wurde und boto für diese Version nicht installiert war.
- Es wurde ein Problem behoben, bei dem YARN-Container und Spark-Anwendungen eine eindeutige Ganglia Round Robin-Datenbankdatei (rrd) generierte, wodurch der erste, der Instance angefügte Datenträger vollständig belegt wurde. Aufgrund dieser Fehlerbehebung wurden YARN-Metriken auf Containerebene und Spark-Metriken auf Anwendungsebene deaktiviert.
- Es wurde ein Problem im Protokoll-Pusher behoben, bei dem alle leeren Protokollordner gelöscht wurden. Das Ergebnis war, dass die Hive-CLI keine Protokolle erstellen konnte, da der Protokoll-Pusher den leeren `user`-Ordner unter `/var/log/hive` entfernte.
- Es wurde ein Problem mit Hive-Importen behoben, da sich auf die Partitionierung auswirkte und beim Import zu einem Fehler führte.
- Es wurde ein Problem behoben, bei dem EMRFS und `s3-dist-cp` Bucket-Namen, die Punkte enthielten, nicht ordnungsgemäß verarbeiteten.
- Es wurde ein Verhalten in EMRFS geändert, damit in Versioning-fähigen Buckets die Markierungsdatei `_$folder$` nicht kontinuierlich erstellt wird. Dies kann zu einer verbesserten Leistung für Versioning-fähige Buckets beitragen.
- Es wurde das Verhalten in EMRFS geändert, sodass Anweisungsdateien nur in Fällen verwendet werden, in denen die clientseitige Verschlüsselung aktiviert ist. Wenn Sie Anweisungsdateien bei Verwendung der clientseitigen Verschlüsselung löschen möchten, können Sie die `emrfs-site.xml`-Eigenschaft, `fs.s3.cse.cryptoStorageMode.deleteInstructionFiles.enabled`, auf "true" festlegen.

- Die YARN-Protokollaggregation wurde geändert, sodass Protokolle am Aggregationsziel zwei Tage lang verwahrt werden. Das Standard-Ziel ist der HDFS-Speicher Ihres Clusters. Wenn Sie diese Dauer ändern möchten, ändern Sie den Wert `yarn.log-aggregation.retain-seconds` mit der Konfigurationsklassifizierung `yarn-site` beim Erstellen Ihres Clusters. Wie immer können Sie Ihre Anwendungsprotokolle beim Erstellen Ihres Clusters mit dem Parameter `log-uri` in Amazon S3 speichern.

## Angewendete Patches

Die folgenden Patches aus Open-Source-Projekten sind in dieser Version enthalten:

- [HIVE-9655](#)
- [HIVE-9183](#)
- [HADOOP-12810](#)

## Version 4.3.0

Veröffentlichungsdatum: 19. Januar 2016

## Features

Die folgenden Funktionen sind in dieser Version verfügbar:

- Upgrade auf Hadoop 2.7.1
- Upgrade auf Spark 1.6.0
- Upgrade auf Ganglia 3.7.2
- Upgrade auf Presto 0.130

Amazon EMR hat einige Änderungen an der Einstellung `spark.dynamicAllocation.enabled` vorgenommen, wenn sie auf „true“ gesetzt wurde. Standardmäßig lautet der Wert „false“.

Wenn die Einstellung auf "true" festgelegt ist, wirkt sich dies auf von der Einstellung `maximizeResourceAllocation` vorgenommene Standardeinstellungen aus:

- Wenn `spark.dynamicAllocation.enabled` auf "true" eingestellt ist, wird `spark.executor.instances` nicht von `maximizeResourceAllocation` festgelegt.

- Die Einstellung `spark.driver.memory` wird nun basierend auf den Instance-Typen im Cluster auf ähnliche Weise konfiguriert, wie `spark.executors.memory` festgelegt wird. Da die Spark-Treiberanwendung jedoch entweder in der Master-Instance oder in einer der Core-Instances ausgeführt werden kann (z. B. im YARN-Client- bzw. -Cluster-Modus), wird die Einstellung `spark.driver.memory` auf der Grundlage des kleineren Instance-Typs in diesen beiden Instance-Gruppen festgelegt.
- Die Einstellung `spark.default.parallelism` wird jetzt auf die doppelte Anzahl von CPU-Kernen für YARN Container festgelegt. In früheren Versionen war dies die Hälfte des Werts.
- Die Berechnungen für den Arbeitsspeicher-Overhead, der für Spark YARN-Prozesse reserviert wurde, wurde präziser angepasst. Dies führt zu einer kleinen Erhöhung des gesamten, verfügbaren Speichers für Spark (d. h. `spark.executor.memory`).

## Bekannte Probleme aus den früheren Versionen, die behoben wurden

- Die YARN-Protokollaggregation ist jetzt standardmäßig aktiviert.
- Es wurde ein Problem behoben, bei dem Protokolle nicht per Push auf den Amazon-S3-Protokoll-Bucket eines Clusters übertragen wurden, wenn die YARN-Protokollaggregation aktiviert war.
- Die Größe von YARN-Containern haben jetzt ein neues Minimum von 32 für alle Knotentypen.
- Es wurde ein Problem mit Ganglia behoben, dass zu übermäßigen Festplatten-E/A-Vorgängen auf dem Master-Knoten in großen Clustern führte.
- Es wurde ein Problem behoben, das verhinderte, dass Anwendungsprotokolle per Push-Verfahren an Amazon S3 übertragen wurden, wenn ein Cluster heruntergefahren wird.
- Es wurde ein Problem in der EMRFS-CLI behoben, das dazu führte, dass einige Befehle fehlschlügen.
- Es wurde ein Problem mit Zeppelin behoben, das verhinderte, dass Abhängigkeiten in die Basisdatei geladen wurden. `SparkContext`
- Es wurde ein Problem behoben, das aus einer Größenanpassung resultierte bei dem Versuch, Instances hinzuzufügen.
- Es wurde ein Problem in Hive behoben, bei dem `CREATE TABLE AS SELECT` übermäßige Listenaufrufe an Amazon S3 erstellt.
- Es wurde ein Problem behoben, bei dem große Cluster nicht ordnungsgemäß bereitgestellt wurden wenn Hue, Oozie und Ganglia installiert sind.
- Es wurde ein Problem in `s3-dist-cp` behoben, bei dem ein Beendigungscode gleich Null zurückgegeben wurde, wenn ein Fehler auftrat.

## Angewendete Patches

Die folgenden Patches aus Open-Source-Projekten sind in dieser Version enthalten:

- [OOZIE-2402](#)
- [HIVE-12502](#)
- [HIVE-10631](#)
- [HIVE-12213](#)
- [HIVE-10559](#)
- [HIVE-12715](#)
- [HIVE-10685](#)

## Version 4.2.0

Veröffentlichungsdatum: 18. November 2015

### Features

Die folgenden Funktionen sind in dieser Version verfügbar:

- Ganglia-Support wurde hinzugefügt.
- Upgrade auf Spark 1.5.2
- Upgrade auf Presto 0.125
- Upgrade auf Oozie 4.2.0
- Upgrade auf Zeppelin 0.5.5
- Auf 1.10.27 AWS SDK for Java aktualisiert

### Bekannte Probleme aus den früheren Versionen, die behoben wurden

- Es wurde ein Problem mit der EMRFS-Befehlszeilenschnittstelle behoben, bei dem der standardmäßige Metadatentabellenname nicht verwendet wurde.
- Es wurde ein Problem behoben, das bei der Verwendung von ORC-gestützten Tabellen in Amazon S3 auftrat.
- Es wurde ein Problem behoben, das bei einem Python-Versionskonflikt in der Spark-Konfiguration auftrat.

- Es wurde ein Problem behoben, bei dem ein YARN-Knotenstatus aufgrund von DNS-Problemen für Cluster in einer VPC keine Daten sendete.
- Es wurde ein Problem behoben, die auftraten, wenn YARN Knoten stilllegte. Das führt dazu, dass Anwendungen hängen blieben oder keine neuen Anwendungen geplant werden konnten.
- Es wurde ein Problem behoben, das auftrat, wenn Cluster mit dem Status `TIMED_OUT_STARTING` beendet wurden.
- Es wurde ein Problem behoben, das auftrat, wenn die EMRFS Scala-Abhängigkeit in andere Builds einbezogen wurde. Die Scala-Abhängigkeit wurde entfernt.

# Anwendungen konfigurieren

Um die Standardkonfigurationen für eine Anwendung zu überschreiben, können Sie ein Konfigurationsobjekt bereitstellen. Sie können entweder eine Syntax-Kurznotation verwenden, um die Konfiguration anzugeben oder Sie können auf die Konfiguration in einer JSON-Datei zu verweisen. Konfigurationsobjekte bestehen aus einer Klassifizierung, Eigenschaften und optionalen verschachtelten Konfigurationen. Die Eigenschaften entsprechen den Anwendungseinstellungen, die Sie ändern möchten. Sie können mehrere Klassifizierungen für mehrere Anwendungen in einem einzigen JSON-Objekt angeben.

## Warning

Die API-Operationen Describe und List von Amazon EMR geben benutzerdefinierte und konfigurierbare Einstellungen, die als Teil der Amazon-EMR-Auftragsabläufe verwendet werden, im Klartext aus. Informationen zur Angabe vertraulicher Informationen wie Kennwörter in diesen Einstellungen finden Sie unter [Vertrauliche Konfigurationsdaten in AWS Secrets Manager speichern](#).

Die verfügbaren Konfigurationsklassifizierungen variieren je nach Amazon-EMR-Version. Eine Liste der Konfigurationsklassifizierungen, die in einer bestimmten Release-Version unterstützt werden, finden Sie auf der Seite für diese Release-Version unter [Informationen über Amazon-EMR-Versionen](#).

Das folgende JSON-Beispieldatei gilt für eine Liste von Konfigurationen.

```
[
  {
    "Classification": "core-site",
    "Properties": {
      "hadoop.security.groups.cache.secs": "250"
    }
  },
  {
    "Classification": "mapred-site",
    "Properties": {
      "mapred.tasktracker.map.tasks.maximum": "2",
      "mapreduce.map.sort.spill.percent": "0.90",
      "mapreduce.tasktracker.reduce.tasks.maximum": "5"
    }
  }
]
```



```
}
]
```

Eine Konfigurationsklassifizierung bezieht sich häufig auf eine anwendungsspezifische Konfigurationsdatei. Die `hive-site`-Klassifizierung ist Einstellungen in der `hive-site.xml`-Konfigurationsdatei für Hive zugeordnet. Eine Ausnahme bildet die veraltete Bootstrap-Aktion `configure-daemons`, mit der Umgebungsparameter wie z. B. `--namenode-heap-size` definiert werden. Optionen wie diese sind in den Klassifizierungen `hadoop-env` und `yarn-env` mit ihren eigenen verschachtelten Exportklassifizierungen zusammengefasst. Verwenden Sie die Export-Unterklassifizierung, wenn eine Klassifizierung auf `env` endet.

Eine weitere Ausnahme ist `s3get`. Sie wurde verwendet, um ein `EncryptionMaterialsProvider`-Kundenobjekt auf jedem Knoten in einem Cluster für die Verwendung bei der clientseitigen Verschlüsselung zu platzieren. Der Klassifizierung `emrfs-site` wurde zu diesem Zweck eine Option hinzugefügt.

Im Folgenden finden Sie ein Beispiel für die `hadoop-env`-Klassifizierung.

```
[
  {
    "Classification": "hadoop-env",
    "Properties": {

    },
    "Configurations": [
      {
        "Classification": "export",
        "Properties": {
          "HADOOP_DATANODE_HEAPSIZE": "2048",
          "HADOOP_NAMENODE_OPTS": "-XX:GCTimeRatio=19"
        },
        "Configurations": [

        ]
      }
    ]
  }
]
```

Im Folgenden finden Sie ein Beispiel für die `yarn-env`-Klassifizierung.

```
[
  {
    "Classification": "yarn-env",
    "Properties": {

    },
    "Configurations": [
      {
        "Classification": "export",
        "Properties": {
          "YARN_RESOURCEMANAGER_OPTS": "-Xdebug -Xrunjdp:transport=dt_socket"
        },
        "Configurations": [

        ]
      }
    ]
  }
]
```

Die folgenden Einstellungen gehören nicht zu einer Konfigurationsdatei, werden aber von Amazon EMR verwendet, um ggf. mehrere Einstellungen in Ihrem Namen zu konfigurieren.

Von Amazon EMR zusammengestellte Einstellungen

Anwendung	Klassifizierung der Versionsbezeichnung	Gültige Eigenschaften	Wann sollte dies verwendet werden?
Spark	spark	maximizeResourceAllocation	Konfigurieren von Executors zur Nutzung der maximalen Ressourcen eines jeden Knotens

Themen

- [Anwendungen bei der Clustererstellung konfigurieren](#)
- [Eine Instance-Gruppe in einem laufenden Cluster neu konfigurieren](#)
- [Vertrauliche Konfigurationsdaten in AWS Secrets Manager speichern](#)
- [Anwendungen für die Verwendung einer bestimmten Java Virtual Machine konfigurieren](#)

## Anwendungen bei der Clustererstellung konfigurieren

Wenn Sie einen Cluster erstellen, können Sie die Standardkonfigurationen für Anwendungen mithilfe der Amazon EMR-Konsole, der AWS Command Line Interface (AWS CLI) oder des AWS SDK überschreiben.

Um die Standardkonfiguration für eine Anwendung zu überschreiben, geben Sie benutzerdefinierte Werte in einer Konfigurationsklassifizierung an. Eine Konfigurationsklassifizierung entspricht einer XML-Konfigurationsdatei für eine Anwendung, z. B. `hive-site.xml`

Konfigurationsklassifizierungen variieren je nach Amazon-EMR-Version. Eine Liste der Konfigurationsklassifizierungen, die in einer bestimmten Release-Version verfügbar sind, finden Sie auf der Release-Detailseite. In diesem Beispiel wird [Amazon-EMR-Version 6.4.0](#) verwendet.

### Geben Sie beim Erstellen eines Clusters eine Konfiguration in der Konsole an

Um eine Konfiguration bereitzustellen, navigieren Sie zur Seite Cluster erstellen und erweitern Sie Softwareeinstellungen. Sie können dann die Konfiguration direkt mittels JSON oder einer Syntax-Kurznotation angeben, dargestellt als abgedunkelter Text in der Konsole. Andernfalls können Sie eine Amazon-S3-URI für eine Datei mit einem `Configurations-JSON-Objekt` bereitstellen.

Um eine Konfiguration für eine Instanzgruppe bereitzustellen, wählen Sie einen Cluster in Ihrer Clusterliste aus und wählen Sie dann die Registerkarte Konfigurationen. Wählen Sie in der Tabelle mit den Instanzgruppenkonfigurationen die zu bearbeitende Instanzgruppe aus und klicken Sie dann auf Neu konfigurieren.

### Geben Sie AWS CLI beim Erstellen eines Clusters eine Konfiguration an, indem Sie

Sie können eine Konfiguration für `create-cluster` bereitstellen, indem Sie den Pfad zu einer JSON-Datei angeben, die lokal oder in Amazon S3 gespeichert ist. Im folgenden Beispiel wird davon ausgegangen, dass Sie bereits erstellte Standardrollen für Amazon EMR verwenden. Wenn Sie die Rollen erstellen müssen, führen Sie zuerst `aws emr create-default-roles` aus.

Wenn sich die Konfiguration im lokalen Verzeichnis befindet, können Sie den folgenden Beispielbefehl verwenden.

```
aws emr create-cluster --use-default-roles --release-label emr-7.1.0 --applications
  Name=Hive \
--instance-type m5.xlarge --instance-count 3 --configurations file:///./
configurations.json
```

Wenn sich Ihre Konfiguration in einem Amazon-S3-Pfad befindet, müssen Sie die folgende Problemumgehung einrichten, bevor Sie den Amazon-S3-Pfad an den `create-cluster`-Befehl übergeben.

```
#!/bin/sh
# Assume the ConfigurationS3Path is not public, and its present in the same AWS account
as the EMR cluster
ConfigurationS3Path="s3://my-bucket/config.json"
# Get a presigned HTTP URL for the s3Path
ConfigurationURL=`aws s3 presign $ConfigurationS3Path --expires-in 300`
# Fetch the presigned URL, and minify the JSON so that it spans only a single line
Configurations=`curl $ConfigurationURL | jq -c .`
aws emr create-cluster --use-default-roles --release-label emr-5.34.0 --instance-type
m5.xlarge --instance-count 2 --applications Name=Hadoop Name=Spark --configurations
$Configurations
```

## Beim Erstellen eines Clusters eine Konfiguration mit dem Java SDK angeben

Der folgende Programmausschnitt zeigt, wie eine Konfiguration mit AWS SDK for Java bereitgestellt wird.

```
Application hive = new Application().withName("Hive");

Map<String,String> hiveProperties = new HashMap<String,String>();
hiveProperties.put("hive.join.emit.interval","1000");
hiveProperties.put("hive.merge.mapfiles","true");

Configuration myHiveConfig = new Configuration()
  .withClassification("hive-site")
  .withProperties(hiveProperties);

RunJobFlowRequest request = new RunJobFlowRequest()
  .withName("Create cluster with ReleaseLabel")
  .withReleaseLabel("emr-5.20.0")
  .withApplications(hive)
```

```
.withConfigurations(myHiveConfig)
.withServiceRole("EMR_DefaultRole")
.withJobFlowRole("EMR_EC2_DefaultRole")
.withInstances(new JobFlowInstancesConfig()
    .withEc2KeyName("myEc2Key")
    .withInstanceCount(3)
    .withKeepJobFlowAliveWhenNoSteps(true)
    .withMasterInstanceType("m4.large")
    .withSlaveInstanceType("m4.large")
);
```

## Eine Instance-Gruppe in einem laufenden Cluster neu konfigurieren

Ab Amazon-EMR-Version 5.21.0 können Sie Cluster-Konfigurationen neu konfigurieren und zusätzliche Konfigurationsklassifikationen für jede Instance-Gruppe in einem ausgeführten Cluster angeben. Dazu können Sie die Amazon EMR-Konsole, das AWS Command Line Interface (AWS CLI) oder das AWS SDK verwenden.

Wenn Sie eine Anwendungskonfiguration für eine Instance-Gruppe in der neuen Amazon-EMR-Konsole aktualisieren, versucht die Konsole, die neue Konfiguration mit der vorhandenen Konfiguration zusammenzuführen, um eine neue, aktive Konfiguration zu erstellen. In dem ungewöhnlichen Fall, dass Amazon EMR die Konfiguration nicht zusammenführen kann, werden Sie von der Konsole benachrichtigt.

Nachdem Sie eine Neukonfigurationsanforderung für eine Instance-Gruppe übermittelt haben, weist Amazon EMR der neuen Konfigurationsspezifikation eine Versionsnummer zu. Sie können die Versionsnummer einer Konfiguration oder den Status einer Instance-Gruppe verfolgen, indem Sie sich die CloudWatch Ereignisse ansehen. Weitere Informationen finden Sie unter [CloudWatch Ereignisse überwachen](#).

### Note

Sie können Clusterkonfigurationen, die bei der Clustererstellung angegeben wurden, nur überschreiben und nicht löschen. Wenn es Unterschiede zwischen der vorhandenen Konfiguration und der von Ihnen bereitgestellten Datei gibt, setzt Amazon EMR manuell geänderte Konfigurationen, z. B. Konfigurationen, die Sie geändert haben, während Sie über SSH mit Ihrem Cluster verbunden waren, auf die Cluster-Standardinstellungen für die angegebene Instance-Gruppe zurück.

# Überlegungen bei der Neukonfiguration einer Instance-Gruppe

## Aktionen zur Neukonfiguration

Wenn Sie eine Rekonfigurationsanfrage über die Amazon EMR-Konsole, das AWS Command Line Interface (AWS CLI) oder das AWS SDK einreichen, überprüft Amazon EMR die vorhandene Cluster-Konfigurationsdatei. Wenn es Unterschiede zwischen der vorhandenen Konfiguration und der von Ihnen bereitgestellten Datei gibt, initiiert Amazon EMR Neukonfigurationsaktionen, startet einige Anwendungen neu und setzt alle manuell geänderten Konfigurationen, wie Konfigurationen, die Sie geändert haben, während Sie über SSH mit Ihrem Cluster verbunden waren, auf die Cluster-Standard Einstellungen für die angegebene Instance-Gruppe zurück.

### Note

Amazon EMR führt bei jeder Neukonfiguration der Instance-Gruppe einige Standardaktionen durch. Diese Standardaktionen können zu Konflikten mit den von Ihnen vorgenommenen Cluster-Anpassungen führen und zu Fehlern bei der Neukonfiguration führen. Weitere Informationen, wie Sie bei fehlgeschlagener Neukonfiguration beheben können, finden Sie unter [Probleme bei der Neukonfiguration von Instance-Gruppen beheben](#).

Amazon EMR leitet auch Neukonfigurationsaktionen für die Konfigurationsklassifizierungen ein, die Sie in Ihrer Anfrage angeben. Eine vollständige Liste dieser Aktionen finden Sie im Abschnitt Konfigurationsklassifizierungen für die Version von Amazon EMR, die Sie verwenden. Zum Beispiel [6.2.0.-Konfigurationsklassifizierungen](#).

### Note

Im Amazon-EMR-Versionshandbuch sind nur Neukonfigurationsaktionen aufgeführt, die mit den Amazon-EMR-Versionen 5.32.0 und 6.2.0 beginnen.

## Betriebsunterbrechung

Amazon EMR führt die Rekonfiguration der Instances in den Instance-Gruppen Core und Task rollierend durch. Zu einem bestimmten Zeitpunkt werden nur 10 % der Instances in einer Instance-Gruppe geändert und neu gestartet. So benötigt der Prozess mehr Zeit, gleichzeitig ist aber die Wahrscheinlichkeit eines Anwendungsausfalls in einem aktiven Cluster geringer.

Um YARN-Aufträge während eines YARN-Neustarts auszuführen, können Sie entweder einen Amazon-EMR-Cluster mit mehreren Hauptknoten erstellen oder `yarn.resourcemanager.recovery.enabled` in Ihrer `yarn-site`-Konfigurationsklassifizierung `true` auf einstellen. [Weitere Informationen zur Verwendung mehrerer Master-Knoten finden Sie unter YARN für hohe Verfügbarkeit. ResourceManager](#)

## Validierung der Anwendung

Amazon EMR überprüft, ob jede Anwendung auf dem Cluster nach dem Neustart der Neukonfiguration ausgeführt wird. Wenn eine Anwendung nicht verfügbar ist, schlägt der gesamte Neukonfigurationsvorgang fehl. Wenn eine Rekonfigurationoperation fehlschlägt, stellt Amazon EMR die Konfigurationsparameter der zuvor funktionierenden Version wieder her.

### Note

Um Fehler bei der Neukonfiguration zu vermeiden, empfehlen wir, nur die Anwendungen auf Ihrem Cluster zu installieren, die Sie verwenden möchten. Wir empfehlen Ihnen außerdem, sicherzustellen, dass alle Clusteranwendungen fehlerfrei sind und ausgeführt werden, bevor Sie eine Rekonfigurationsanfrage einreichen.

## Arten der Neukonfiguration

Sie können eine Instance-Gruppe auf zwei Arten neu konfigurieren:

- **Überschreiben.** Standardmethode für die Rekonfiguration und die einzige, die in Amazon-EMR-Versionen vor 5.35.0 und 6.6.0 verfügbar ist. Diese Rekonfigurationsmethode überschreibt wahllos alle Dateien im Cluster mit dem neu eingereichten Konfigurationssatz. Die Methode löscht alle Änderungen an Konfigurationsdateien, die außerhalb der Rekonfigurations-API vorgenommen wurden.
- **Mischen.** Die Rekonfigurationsmethode wird für die Amazon-EMR-Versionen 5.35.0 und 6.6.0 und höher unterstützt, außer in der Amazon-EMR-Konsole, wo sie von keiner Version unterstützt wird. Diese Rekonfigurationsmethode führt die neu eingereichten Konfigurationen mit Konfigurationen zusammen, die bereits auf dem Cluster vorhanden sind. Diese Option fügt nur die neuen Konfigurationen hinzu oder ändert sie, die Sie einreichen. Sie behält bestehende Konfigurationen bei.

 Note

Amazon EMR überschreibt weiterhin einige wichtige Hadoop-Konfigurationen, die es benötigt, um sicherzustellen, dass der Service ordnungsgemäß ausgeführt wird.

## Einschränkungen

Beachten Sie bei der Neukonfiguration einer Instance-Gruppe in einem laufenden Cluster die folgenden Einschränkungen:

- Anwendungen, die nicht von Yarn stammen, können beim Neustart fehlschlagen oder Clusterprobleme verursachen, insbesondere wenn die Anwendungen nicht richtig konfiguriert sind. Bei Clustern, die sich der maximalen Speicher- und CPU-Auslastung nähern, können nach dem Neustart Probleme auftreten. Dies gilt insbesondere für die Master-Instance-Gruppe.
- Sie können keine Rekonfigurationsanfrage stellen, wenn die Größe einer Instance-Gruppe geändert wird. Wenn eine Rekonfiguration initiiert wird, während die Größe einer Instance-Gruppe geändert wird, kann die Rekonfiguration erst durchgeführt werden, nachdem die Größenänderung abgeschlossen wurde (und umgekehrt).
- Nachdem eine Instance-Gruppe rekonfiguriert wurde, startet Amazon EMR die Anwendungen, um die neuen Konfigurationen zu aktivieren. Auftragsfehler und anderes unerwartetes Anwendungsverhalten kann auftreten, wenn die Anwendungen während der Rekonfiguration genutzt werden.
- Wenn eine Rekonfiguration für eine Instance-Gruppe fehlschlägt, stellt Amazon EMR die Konfigurationsparameter der zuvor funktionierenden Version wieder her. Schlägt auch diese Wiederherstellung fehl, müssen Sie eine neue `ModifyInstanceGroup`-Anforderung zur Wiederherstellung der Instance-Gruppe aus dem Status `SUSPENDED` senden.
- Rekonfigurationsanforderungen für Phoenix-Konfigurationsklassifizierungen werden nur in Amazon EMR ab Version 5.23.0, nicht jedoch in den Amazon-EMR-Versionen 5.21.0 und 5.22.0 unterstützt.
- Rekonfigurationsanfragen für HBase-Konfigurationsklassifizierungen werden nur in Amazon-EMR-Version 5.30.0 und höher unterstützt und werden in den Amazon-EMR-Versionen 5.23.0 bis 5.29.0 nicht unterstützt.
- Amazon EMR unterstützt Anfragen zur Neukonfiguration von Anwendungen auf einem Amazon-EMR-Cluster mit mehreren Primärknoten nur in Amazon-EMR-Versionen 5.27.0 und höher.



- Die Neukonfiguration der `hdfs-encryption-zones`-Klassifizierung oder einer der Hadoop KMS-Konfigurationsklassifizierungen auf einem ausgeführten Cluster wird in Amazon EMR nicht unterstützt.
- Amazon EMR unterstützt derzeit bestimmte Rekonfigurationsanfragen für den Capacity Scheduler nicht, die einen Neustart des YARN erfordern. ResourceManager Beispielsweise können Sie eine Warteschlange nicht vollständig entfernen.

## Eine Instance-Gruppe in der Konsole neu konfigurieren

### Note

Die Amazon-EMR-Konsole unterstützt keine Rekonfigurationen vom Typ Merge.

1. [Öffnen Sie die Amazon EMR-Konsole unter `https://console.aws.amazon.com/emr`](https://console.aws.amazon.com/emr)
2. Wählen Sie in der Cluster-Liste den aktiven Cluster, der umkonfiguriert werden soll, unter Name.
3. Öffnen Sie die Seite mit den Cluster-Details für den Cluster und wechseln Sie zur Registerkarte Konfigurationen.
4. Wählen Sie in der Dropdown-Liste Filter die Instance-Gruppe aus, die rekonfiguriert werden soll.
5. Wählen Sie in der Dropdown-Liste Rekonfigurieren die Option In Tabelle bearbeiten oder In JSON-Datei bearbeiten.
  - In Tabelle bearbeiten: Bearbeiten Sie in der Tabelle der Konfigurationsklassifizierungen Eigenschaft und Wert vorhandener Konfigurationen oder wählen Sie Konfiguration hinzufügen, um für weitere Konfigurationsklassifizierungen anzugeben.
  - In JSON-Datei bearbeiten – Geben Sie die Konfiguration direkt in JSON ein oder verwenden Sie die Kurzschriftsyntax (demonstriert im Schattentext). Stellen Sie andernfalls eine Amazon-S3-URI für eine Datei mit einem JSON-Configurations-Objekt bereit.

### Note

Die Spalte Quelle in der Tabelle der Konfigurationsklassifizierungen zeigt, ob die Konfiguration beim Erstellen eines Clusters oder beim Angeben zusätzlicher Konfigurationen für diese Instance-Gruppe bereitgestellt wird. Sie können Konfigurationen für eine Instance-Gruppe aus beiden Quellen bearbeiten. Sie können die

ursprünglichen Cluster-Konfigurationen nicht löschen, aber sie für eine Instance-Gruppe überschreiben.

Sie können außerdem verschachtelte Konfigurationsklassifizierungen direkt in der Tabelle hinzufügen oder bearbeiten. Um beispielsweise eine zusätzliche `export-` Unterklassifizierung von `hadoop-env` bereitzustellen, fügen Sie eine `hadoop.export-` Konfigurationsklassifizierung in die Tabelle ein. Anschließend stellen Sie eine Eigenschaft samt Wert für diese Klassifizierung bereit.

6. (Optional) Wählen Sie `Apply this configuration to all active instance groups` (Diese Konfiguration allen aktiven Instance-Gruppen zuweisen).
7. Speichern Sie die Änderungen.

## Eine Instance-Gruppe mithilfe der CLI rekonfigurieren

Verwenden Sie den Befehl `modify-instance-groups`, um eine neue Konfiguration für eine Instance-Gruppe in einem laufenden Cluster anzugeben.

### Note

Ersetzen Sie in den folgenden Beispielen `<j-2AL4XXXXXX5T9>` durch die Cluster-ID und `<ig-1xxxxxxx9>` durch die ID der Instance-Gruppe.

### Example – Eine Konfiguration für eine Instance-Gruppe ersetzen

Das folgende Beispiel verweist auf eine JSON-Konfigurationsdatei, die aufgerufen wird `instanceGroups.json`, um die Eigenschaft des NodeManager YARN-Festplattenintegritätsprüfers für eine Instanzgruppe zu bearbeiten.

1. Bereiten Sie Ihre Konfigurationsklassifizierung vor und speichern Sie sie als `instanceGroups.json` in demselben Verzeichnis, in dem Sie den Befehl ausführen werden.

```
[
  {
    "InstanceGroupId": "<ig-1xxxxxxx9>",
    "Configurations": [
      {
        "Classification": "yarn-site",
```

```

        "Properties":{
            "yarn.nodemanager.disk-health-checker.enable":"true",
            "yarn.nodemanager.disk-health-checker.max-disk-utilization-per-disk-
percentage":"100.0"
        },
        "Configurations":[]
    }
]
}
]

```

2. Führen Sie den folgenden Befehl aus.

```
aws emr modify-instance-groups --cluster-id <j-2AL4XXXXXX5T9> \
--instance-groups file://instanceGroups.json
```

### Example – Einer Instance-Gruppe eine Konfiguration hinzufügen

Wenn Sie einer Instance-Gruppe eine Konfiguration hinzufügen möchten, müssen Sie alle zuvor angegebenen Konfigurationen für diese Instance-Gruppe in Ihre neue `ModifyInstanceGroup`-Anforderung aufnehmen. Andernfalls werden die zuvor für die angegebenen Konfigurationen entfernt.

Das folgende Beispiel fügt eine Eigenschaft für den NodeManager virtuellen YARN-Speicherprüfer hinzu. Die Konfiguration enthält auch zuvor angegebene Werte für den NodeManager YARN-Festplattenintegritätsprüfer, sodass die Werte nicht überschrieben werden.

1. Bereiten Sie den folgenden Inhalt in `instanceGroups.json` vor und speichern Sie ihn im selben Verzeichnis, in dem Sie den Befehl ausführen werden.

```

[
  {
    "InstanceId":"<ig-1xxxxxxx9>",
    "Configurations":[
      {
        "Classification":"yarn-site",
        "Properties":{
          "yarn.nodemanager.disk-health-checker.enable":"true",
          "yarn.nodemanager.disk-health-checker.max-disk-utilization-per-disk-
percentage":"100.0",

```

```

        "yarn.nodemanager.vmem-check-enabled":"true",
        "yarn.nodemanager.vmem-pmem-ratio":"3.0"
    },
    "Configurations":[]
}
]
}
]

```

2. Führen Sie den folgenden Befehl aus.

```
aws emr modify-instance-groups --cluster-id <j-2AL4XXXXXX5T9> \
--instance-groups file://instanceGroups.json
```

Example – Einer Instance-Gruppe mit Rekonfiguration vom Typ Zusammenführen eine Konfiguration hinzufügen

Wenn Sie die Standardmethode Rekonfiguration Überschreiben verwenden möchten, um eine Konfiguration hinzuzufügen, müssen Sie alle zuvor angegebenen Konfigurationen für diese Instance-Gruppe in Ihre neue ModifyInstanceGroup-Anfrage aufnehmen. Andernfalls entfernt Überschreiben die Konfigurationen, die Sie zuvor angegeben haben. Bei der Neukonfiguration mit Zusammenführen müssen Sie dies nicht tun. Stattdessen müssen Sie sicherstellen, dass Ihre Anfrage nur die neuen Konfigurationen enthält.

Das folgende Beispiel fügt eine Eigenschaft für den NodeManager virtuellen YARN-Speicherprüfer hinzu. Da es sich um eine Rekonfiguration vom Typ Merge handelt, werden zuvor angegebene Werte für die NodeManager YARN-Festplattenintegritätsprüfung nicht überschrieben.

1. Bereiten Sie den folgenden Inhalt in `instanceGroups.json` vor und speichern Sie ihn im selben Verzeichnis, in dem Sie den Befehl ausführen werden.

```

[
  {"InstanceId":"<ig-1xxxxxxx9>",
    "ReconfigurationType":"MERGE",
    "Configurations":[
      {"Classification":"yarn-site",
        "Properties":{
          "yarn.nodemanager.vmem-check-enabled":"true",
          "yarn.nodemanager.vmem-pmem-ratio":"3.0"
        }
      }
    ]
  }
]

```

```

    },
    "Configurations":[]
  }
]
}
]

```

2. Führen Sie den folgenden Befehl aus.

```
aws emr modify-instance-groups --cluster-id <j-2AL4XXXXXX5T9> \
--instance-groups file://instanceGroups.json
```

### Example – Eine Konfiguration für eine Instance-Gruppe löschen

Um eine Konfiguration für eine Instance-Gruppe zu löschen, senden Sie eine neue Neukonfigurationsanforderung, die die vorherige Konfiguration ausschließt.

#### Note

Sie können die ursprüngliche Cluster-Konfiguration nur überschreiben. Sie kann nicht gelöscht werden.

Um beispielsweise die Konfiguration für die NodeManager YARN-Festplattenintegritätsprüfung aus dem vorherigen Beispiel zu löschen, reichen Sie eine neue `instanceGroups.json` mit dem folgenden Inhalt ein.

```

[
  {
    "InstanceId": "<ig-1xxxxxxx9>",
    "Configurations": [
      {
        "Classification": "yarn-site",
        "Properties": {
          "yarn.nodemanager.vmem-check-enabled": "true",
          "yarn.nodemanager.vmem-pmem-ratio": "3.0"
        }
      },
      "Configurations": []
    ]
  }
]

```

```

    ]
  }
]

```

### Note

Um alle Konfigurationen der letzten Rekonfigurationsanforderung zu löschen, senden Sie eine Rekonfigurationsanforderung mit einem leeren Konfigurations-Array. Zum Beispiel

```

[
  {
    "InstanceId": "<ig-1xxxxxxx9>",
    "Configurations": []
  }
]

```

Example – Eine Instance-Gruppe in einer Anfrage neu konfigurieren oder die Größe ändern

Das folgende JSON-Beispiel zeigt, wie eine Instance-Gruppe in derselben Anfrage neu konfiguriert und in der Größe geändert wird.

```

[
  {
    "InstanceId": "<ig-1xxxxxxx9>",
    "InstanceCount": 5,
    "EC2InstanceIdsToTerminate": ["i-123"],
    "ForceShutdown": true,
    "ShrinkPolicy": {
      "DecommissionTimeout": 10,
      "InstanceResizePolicy": {
        "InstancesToTerminate": ["i-123"],
        "InstancesToProtect": ["i-345"],
        "InstanceTerminationTimeout": 20
      }
    },
    "Configurations": [
      {
        "Classification": "yarn-site",

```

```

        "Configurations": [],
        "Properties": {
            "yarn.nodemanager.disk-health-checker.enable": "true",
            "yarn.nodemanager.disk-health-checker.max-disk-utilization-per-disk-
percentage": "100.0"
        }
    ]
}
]

```

## Eine Instance-Gruppe mithilfe des Java SDK neu konfigurieren

### Note

Ersetzen Sie in den folgenden Beispielen `<j-2AL4XXXXXX5T9>` durch die Cluster-ID und `<ig-1xxxxxxx9>` durch die ID der Instance-Gruppe.

Der folgende Codeausschnitt bietet eine neue Konfiguration für eine Instance-Gruppe unter Verwendung von AWS SDK for Java.

```

AWSCredentials credentials = new BasicAWSCredentials("access-key", "secret-key");
AmazonElasticMapReduce emr = new AmazonElasticMapReduceClient(credentials);

Map<String,String> hiveProperties = new HashMap<String,String>();
hiveProperties.put("hive.join.emit.interval","1000");
hiveProperties.put("hive.merge.mapfiles","true");

Configuration configuration = new Configuration()
    .withClassification("hive-site")
    .withProperties(hiveProperties);

InstanceGroupModifyConfig igConfig = new InstanceGroupModifyConfig()
    .withInstanceId("<ig-1xxxxxxx9>")
    .withReconfigurationType("MERGE");
    .withConfigurations(configuration);

ModifyInstanceGroupsRequest migRequest = new ModifyInstanceGroupsRequest()
    .withClusterId("<j-2AL4XXXXXX5T9>")

```

```
.withInstanceGroups(igConfig);  
  
emr.modifyInstanceGroups(migRequest);
```

Der folgende Codeausschnitt löscht eine zuvor angegebene Konfiguration für eine Instance-Gruppe, indem ein leeres Array von Konfigurationen bereitgestellt wird.

```
List<Configuration> configurations = new ArrayList<Configuration>();  
  
InstanceGroupModifyConfig igConfig = new InstanceGroupModifyConfig()  
    .withInstanceId("<ig-1xxxxxxx9>")  
    .withConfigurations(configurations);  
  
ModifyInstanceGroupsRequest migRequest = new ModifyInstanceGroupsRequest()  
    .withClusterId("<j-2AL4XXXXXX5T9>")  
    .withInstanceGroups(igConfig);  
  
emr.modifyInstanceGroups(migRequest);
```

## Probleme bei der Neukonfiguration von Instance-Gruppen beheben

Wenn der Neukonfigurationsprozess für eine Instance-Gruppe fehlschlägt, macht Amazon EMR die Neukonfiguration rückgängig und protokolliert eine Fehlermeldung mithilfe eines Amazon-Ereignisses. CloudWatch Das Ereignis bietet eine kurze Zusammenfassung des Neukonfigurationsfehlers. Es listet die Instances auf, bei denen die Rekonfiguration fehlgeschlagen ist, sowie die entsprechenden Fehlermeldungen. Das Folgende ist ein Beispiel für eine Fehlermeldung.

```
The reconfiguration operation for instance group ig-1xxxxxxx9 in Amazon EMR  
cluster j-2AL4XXXXXX5T9 (ExampleClusterName)  
failed at 2021-01-01 00:00 UTC and took 2 minutes to fail. Failed configuration version  
is example12345.  
Failure message: Instance i-xxxxxxx1, i-xxxxxxx2, i-xxxxxxx3 failed with message "This  
is an example failure message".
```

Weitere Daten zu einem Fehler bei der Rekonfiguration können Sie in den Protokollen zur Knotenbereitstellung nachlesen. Dies ist besonders nützlich, wenn Sie eine Nachricht wie die folgende erhalten.



```
i-xxxxxxx1 failed with message "Unable to complete transaction and some changes were applied."
```

## On the node

So greifen Sie auf die Bereitstellungsprotokolle von Knoten zu, indem Sie eine Verbindung zu einem Knoten herstellen

1. Verwenden Sie SSH, um eine Verbindung zu dem Knoten herzustellen, auf dem die Neukonfiguration fehlgeschlagen ist. Anweisungen finden Sie unter [Herstellen einer Verbindung mit Ihrer Linux-Instance](#) im Amazon-EC2-Benutzerhandbuch für Linux-Instances.
2. Navigieren Sie zum folgenden Verzeichnis, das die Protokolldateien für die Knotenbereitstellung enthält.

```
/mnt/var/log/provision-node/
```

3. Öffnen Sie das `reports`-Unterverzeichnis und suchen Sie nach dem Knotenbereitstellungsbericht für Ihre Neukonfiguration. Das `reports`-Verzeichnis organisiert Protokolle nach Versionsnummer der Rekonfiguration, UUID (Universally Unique Identifier), IP-Adresse der Amazon-EC2-Instance und Zeitstempel. Jeder Bericht ist eine komprimierte YAML-Datei, die detaillierte Informationen über den Rekonfigurationsprozess enthält.

Im Folgenden finden Sie ein Beispiel für einen Berichts-Dateiname und einen Pfad.

```
/reports/2/ca598xxx-cxxx-4xxx-bxxx-6dbxxxxxxxxxx/ip-10-73-xxx-xxx.ec2.internal/202104061715.yaml.gz
```

4. Sie können einen Bericht mit einem Dateibetrachter wie `zless` im folgenden Beispiel untersuchen.

```
zless 202104061715.yaml.gz
```

## Amazon S3

So greifen Sie mit Amazon S3 auf Knotenbereitstellungsprotokolle zu

1. Melden Sie sich bei der Amazon S3 S3-Konsole an AWS Management Console und öffnen Sie sie unter <https://console.aws.amazon.com/s3/>.

2. Öffnen Sie den Amazon-S3-Bucket, den Sie angegeben haben, als Sie den Cluster für die Archivierung von Protokolldateien in konfiguriert haben.
3. Navigieren Sie zum folgenden Ordner, der die Protokolldateien für die Knotenbereitstellung enthält:

```
DOC-EXAMPLE-BUCKET/elasticmapreduce/<cluster id>/node/<instance id>/provision-
node/
```

4. Öffnen Sie den `reports`-Ordner und suchen Sie nach dem Bericht zur Knotenbereitstellung für Ihre Rekonfiguration. Der `reports`-Ordner organisiert die Protokolle nach Versionsnummer der Rekonfiguration, UUID (Universally Unique Identifier), IP-Adresse der Amazon EC2 Instance und Zeitstempel. Jeder Bericht ist eine komprimierte YAML-Datei, die detaillierte Informationen über den Rekonfigurationsprozess enthält.

Im Folgenden finden Sie ein Beispiel für einen Berichts-Dateiname und einen Pfad.

```
/reports/2/ca598xxx-cxxx-4xxx-bxxx-6dbxxxxxxxxx/ip-10-73-xxx-
xxx.ec2.internal/202104061715.yaml.gz
```

5. Um eine Protokolldatei anzuzeigen, können Sie sie als Textdatei von Amazon S3 auf Ihren lokalen Computer herunterladen. Anweisungen finden Sie unter [Objekt herunterladen](#).

Jede Protokolldatei enthält einen detaillierten Bereitstellungsbericht für die zugehörige Rekonfiguration. Um Informationen zu Fehlermeldungen zu finden, können Sie nach der `err`-Protokollebene eines Berichts suchen. Das Berichtsformat hängt von der Amazon-EMR-Version auf Ihrem Cluster ab.

Das folgende Beispiel zeigt Fehlerinformationen für Amazon-EMR-Release-Versionen vor 5.32.0 und 6.2.0.

```
- !ruby/object:Puppet::Util::Log
  level: !ruby/sym err
  tags:
    - err
  message: "Example detailed error message."
  source: Puppet
  time: 2021-01-01 00:00:00.000000 +00:00
```

Die Amazon-EMR-Release-Versionen 5.32.0 und 6.2.0 und höher verwenden stattdessen das folgende Format.

```
- level: err
  message: 'Example detailed error message.'
  source: Puppet
  tags:
  - err
  time: '2021-01-01 00:00:00.000000 +00:00'
  file:
  line:
```

## Vertrauliche Konfigurationsdaten in AWS Secrets Manager speichern

Amazon EMR beschreibt und listet API-Vorgänge auf, die benutzerdefinierte Konfigurationsdaten (wie `DescribeCluster` und `ListInstanceGroups`) ausgeben, und zwar in Klartext. Amazon EMR ist integriert, AWS Secrets Manager sodass Sie Ihre Daten in Secrets Manager speichern und den geheimen ARN in Ihren Konfigurationen verwenden können. Auf diese Weise geben Sie keine vertraulichen Konfigurationsdaten in Amazon EMR im Klartext weiter und stellen sie externen APIs zur Verfügung. Wenn Sie angeben, dass ein Schlüssel-Wert-Paar einen ARN für ein in Secrets Manager gespeichertes Geheimnis enthält, ruft Amazon EMR dieses Geheimnis ab, wenn es Konfigurationsdaten an den Cluster sendet. Amazon EMR sendet die Anmerkung nicht, wenn externe APIs zur Anzeige der Konfiguration verwendet werden.

### Ein Secret erstellen

Um ein Geheimnis zu erstellen, befolgen Sie die Schritte unter [AWS Secrets Manager -Geheimnis erstellen](#) im AWS Secrets Manager -Benutzerhandbuch. In Schritt 3 müssen Sie das Klartext-Feld auswählen, um Ihren sensiblen Wert einzugeben.

Beachten Sie, dass Secrets Manager zwar zulässt, dass ein Geheimnis bis zu 65 536 Byte enthält, Amazon EMR jedoch die kombinierte Länge des Eigenschaftsschlüssels (ohne Anmerkung) und des abgerufenen geheimen Werts auf 1 024 Zeichen begrenzt.

### Gewähren Sie Amazon EMR Zugriff, um das Geheimnis abzurufen

Amazon EMR verwendet eine IAM-Servicerolle, um Cluster für Sie bereitzustellen und zu verwalten. Die Servicerolle für Amazon-EMR-Rolle definiert die zulässigen Aktionen für bei der Bereitstellung

von Ressourcen und der Ausführung von Service-Level-Aufgaben, die nicht im Kontext einer EC2-Instance innerhalb eines Clusters ausgeführt werden. Weitere Informationen zu Servicerollen finden Sie unter [Servicerolle für Amazon EMR \(EMR-Rolle\)](#) und [IAM-Rollen anpassen](#).

Damit Amazon EMR den geheimen Wert aus Secrets Manager abrufen kann, fügen Sie Ihrer Amazon-EMR-Rolle die folgende Richtlinienerklärung hinzu, wenn Sie Ihren Cluster starten.

```
{
  "Sid": "AllowSecretsRetrieval",
  "Effect": "Allow",
  "Action": "secretsmanager:GetSecretValue",
  "Resource": [
    "arn:aws:secretsmanager:<region>:<aws-account-id>:secret:<secret-name>"
  ]
}
```

Wenn Sie den geheimen Schlüssel mit einem vom Kunden verwalteten Schlüssel erstellen AWS KMS key, müssen Sie der Amazon EMR-Rolle auch eine `kms:Decrypt` Berechtigung für den von Ihnen verwendeten Schlüssel hinzufügen. Weitere Informationen finden Sie unter [Authentifizierung und Zugriffskontrolle für AWS Secrets Manager](#) im AWS Secrets Manager -Benutzerhandbuch.

## Das Secret in einer Konfigurationsklassifizierung verwenden

Sie können die `EMR.secret@`-Anmerkung zu jeder Konfigurationseigenschaft hinzufügen, um anzugeben, dass ihr Schlüssel-Wert-Paar einen ARN für ein in Secrets Manager gespeichertes Secret enthält.

Das folgende Beispiel veranschaulicht, wie Sie eine geheime ARN in einer Konfigurationsklassifizierung angeben:

```
{
  "Classification": "core-site",
  "Properties": {
    "presto.s3.access-key": "<sensitive-access-key>",
    "EMR.secret@presto.s3.secret-key": "arn:aws:secretsmanager:<region>:<aws-account-id>:secret:<secret-name>"
  }
}
```

Wenn Sie Ihren Cluster erstellen und Ihre kommentierte Konfiguration einreichen, validiert Amazon EMR die Konfigurationseigenschaften. Wenn Ihre Konfiguration gültig ist, entfernt Amazon EMR die

Anmerkung aus der Konfiguration und ruft das Secret von Secrets Manager ab, um die eigentliche Konfiguration zu erstellen, bevor sie auf den Cluster angewendet wird:

```
{
  "Classification":"core-site",
  "Properties":{
    "presto.s3.access-key":"<sensitive-access-key>",
    "presto.s3.secret-key":"<my-secret-key-retrieved-from-Secrets-Manager>"
  }
}
```

Wenn Sie eine Aktion wie `DescribeCluster` aufrufen, gibt Amazon EMR die aktuelle Anwendungskonfiguration auf dem Cluster zurück. Wenn eine Anwendungskonfigurationseigenschaft so gekennzeichnet ist, dass sie einen geheimen ARN enthält, enthält die vom `DescribeCluster`-Aufruf zurückgegebene Anwendungskonfiguration den ARN und nicht den geheimen Wert. Dadurch wird sichergestellt, dass der geheime Wert nur auf dem Cluster sichtbar ist:

```
{
  "Classification":"core-site",
  "Properties":{
    "presto.s3.access-key":"<sensitive-access-key>",
    "presto.s3.secret-key":"arn:aws:secretsmanager:<region>:<aws-account-id>:secret:<secret-name>"
  }
}
```

## Den Secret-Wert aktualisieren

Amazon EMR ruft den geheimen Wert aus einer mit Anmerkungen versehenen Konfiguration ab, wenn die angehängte Instance-Gruppe gestartet, neu konfiguriert oder ihre Größe geändert wird. Sie können Secrets Manager verwenden, um den Wert eines Secrets zu ändern, das in der Konfiguration eines laufenden Clusters verwendet wird. Wenn Sie dies tun, können Sie eine Rekonfigurationsanfrage an jede Instance-Gruppe senden, die den aktualisierten Wert erhalten soll. Weitere Informationen zur Neukonfiguration einer Instance-Gruppe und zu den Dingen, die Sie dabei beachten sollten, finden Sie unter [Eine Instance-Gruppe in einem laufenden Cluster neu konfigurieren](#).

# Anwendungen für die Verwendung einer bestimmten Java Virtual Machine konfigurieren

Amazon-EMR-Versionen haben unterschiedliche Standardversionen von Java Virtual Machine (JVM). Auf dieser Seite wird die JVM-Unterstützung für verschiedene Versionen und Anwendungen erklärt.

## Überlegungen

Weitere Informationen über die unterstützten Java-Versionen für Anwendungen finden Sie in den [Amazon-EMR-Versionshinweisen](#).

- Amazon EMR unterstützt nur die Ausführung einer Laufzeit-Version in einem Cluster und nicht die Ausführung verschiedener Knoten oder Anwendungen auf verschiedenen Laufzeit-Versionen auf demselben Cluster.
- Für Amazon EMR 7.x ist Java 17 die standardmäßige Java Virtual Machine (JVM) für Anwendungen, die Java 17 unterstützen, mit Ausnahme von Apache Livy. Weitere Informationen über die unterstützten JDK-Versionen für Anwendungen finden Sie in den entsprechenden Versionsseiten in den Amazon-EMR-Versionshinweisen.
- Ab Amazon EMR 7.1.0 unterstützt Flink Java 17 und ist standardmäßig auf Java 17 eingestellt. Um eine andere Version der Java-Runtime zu verwenden, überschreiben Sie die Einstellungen in `flink-conf`. Weitere Informationen zur Konfiguration von Flink für die Verwendung von Java 8 oder Java 11 finden [Sie unter Konfigurieren von Flink für die Ausführung mit Java 11](#).
- Für Amazon EMR 5.x und 6.x ist Java 8 die standardmäßige Java Virtual Machine (JVM).
  - Für Amazon-EMR-Versionen 6.12.0 und höher unterstützen einige Anwendungen auch Java 11 und 17.
  - Für Amazon-EMR-Versionen 6.9.0 und höher unterstützt Trino standardmäßig Java 17. Weitere Informationen über Java 17 mit Trino finden Sie unter [Trino-Updates für Java 17](#) im Trino-Blog.

Beachten Sie bei der Auswahl Ihrer Laufzeit-Version die folgenden anwendungsspezifischen Überlegungen:

### Anwendungsspezifische Java-Konfigurationshinweise

Anwendung	Hinweise zur Java-Konfigurationen
Spark	Um Spark mit einer nicht standardmäßigen Java-Version auszuführen, müssen Sie sowohl

Anwendung	Hinweise zur Java-Konfigurationen
	<p>Spark als auch Hadoop konfigurieren. Beispiele finden Sie unter <a href="#">Die JVM überschreiben</a>.</p> <ul style="list-style-type: none"><li>• Konfigurieren Sie <code>JAVA_HOME</code> in <code>spark-env</code>, um die Java-Laufzeit der primären Instanceprozesse zu aktualisieren. Zum Beispiel <code>spark-submit</code>, <code>spark-shell</code> und Spark History Server.</li><li>• Ändern Sie die Hadoop-Konfiguration, um die Java-Laufzeit der Spark-Executors und des YARN ApplicationMaster zu aktualisieren</li></ul>
Spark RAPIDS	Sie können RAPIDS mit der konfigurierten Java-Version für Spark ausführen.
Iceberg	Sie können Iceberg mit der konfigurierten Java-Version der Anwendung ausführen, die es verwendet.
Delta	Sie können Delta mit der konfigurierten Java-Version der Anwendung ausführen, die es verwendet.
Hudi	Sie können Hudi mit der konfigurierten Java-Version der Anwendung ausführen, die es verwendet.
Hadoop	Um die JVM für Hadoop zu aktualisieren, ändern Sie <code>hadoop-env</code> . Beispiele finden Sie unter <a href="#">Die JVM überschreiben</a> .

Anwendung	Hinweise zur Java-Konfigurationen
Hive	Um die Java-Version für Hive auf 11 oder 17 festzulegen, konfigurieren Sie die Hadoop JVM-Einstellung auf die Java-Version, die Sie verwenden möchten.
HBase	Um die JVM für HBase zu aktualisieren, ändern Sie <code>hbase-env</code> . Standardmäßig legt Amazon EMR die HBase-JVM auf der Grundlage der JVM-Konfiguration für Hadoop fest, sofern Sie die Einstellungen in <code>hbase-env</code> nicht überschreiben. Beispiele finden Sie unter <a href="#">Die JVM überschreiben</a> .
Flink	Um die JVM für Flink zu aktualisieren, ändern Sie <code>flink-conf</code> . Standardmäßig legt Amazon EMR die Flink-JVM auf der Grundlage der JVM-Konfiguration für Hadoop fest, sofern Sie die Einstellungen in <code>flink-conf</code> nicht überschreiben. Weitere Informationen finden Sie unter <a href="#">Flink für die Ausführung mit Java 11 konfigurieren</a> .
Oozie	Um Oozie für die Ausführung unter Java 11 oder 17 zu konfigurieren, konfigurieren Sie Oozie Server, den Oozie LauncherAM Launcher AM und ändern Sie Ihre clientseitigen ausführbaren Dateien und Auftragskonfigurationen. Sie können <code>EmbeddedOozieServer</code> auch für die Ausführung unter Java 17 konfigurieren. Weitere Informationen finden Sie unter <a href="#">Die Java-Version für Oozie konfigurieren</a> .
Pig	Pig unterstützt nur Java 8. Sie können Java 11 oder 17 nicht mit Hadoop verwenden und Pig auf demselben Cluster ausführen.



## Die JVM überschreiben

Um die JVM-Einstellung für eine Amazon-EMR-Version zu überschreiben – z. B. um Java 17 mit einem Cluster zu verwenden, der Amazon-EMR-Version 6.12.0 verwendet – geben Sie die JAVA\_HOME-Einstellung an die Umgebungsklassifizierung an, die *application*-env für alle Anwendungen außer Flink ist. Für Flink lautet die Umgebungsklassifizierung `flink-conf`. Schritte zur Konfiguration der Java-Laufzeit mit Flink finden Sie unter [Flink für die Ausführung mit Java 11 konfigurieren](#).

### Themen

- [Die JVM-Einstellung mit Apache Spark überschreiben](#)
- [Die JVM-Einstellung mit Apache HBase überschreiben](#)
- [Die JVM-Einstellung mit Apache Hadoop und Hive überschreiben](#)

## Die JVM-Einstellung mit Apache Spark überschreiben

Wenn Sie Spark mit Amazon-EMR-Versionen 6.12 und höher verwenden und einen Treiber für die Einreichung im Cluster-Modus schreiben, verwendet der Treiber Java 8, aber Sie können die Umgebung so einstellen, dass die Executors Java 11 oder 17 verwenden. Und wenn Sie Spark mit Amazon-EMR-Versionen unter 5.x verwenden und einen Treiber für die Einreichung im Cluster-Modus schreiben, verwendet der Treiber Java 7. Die Umgebung kann dann aber so eingestellt werden, dass die Executors Java 8 verwenden.

Um die JVM für Spark zu überschreiben, empfehlen wir, dass Sie sowohl die Hadoop- als auch die Spark-Klassifizierung festlegen.

```
{
  "Classification": "hadoop-env",
    "Configurations": [
      {
        "Classification": "export",
          "Configurations": [],
          "Properties": {
            "JAVA_HOME": "/usr/lib/jvm/java-1.8.0"
          }
        }
      ],
      "Properties": {}
    },
  }
```

```

{
  "Classification": "spark-env",
    "Configurations": [
      {
        "Classification": "export",
          "Configurations": [],
          "Properties": {
            "JAVA_HOME": "/usr/lib/jvm/java-1.8.0"
          }
        }
      ],
      "Properties": {}
    }
  }

```

## Die JVM-Einstellung mit Apache HBase überschreiben

Um HBase für die Verwendung von Java 11 zu konfigurieren, können Sie beim Starten des Clusters die folgende Konfiguration festlegen.

```

[
  {
    "Classification": "hbase-env",
      "Configurations": [
        {
          "Classification": "export",
            "Configurations": [],
            "Properties": {
              "JAVA_HOME": "/usr/lib/jvm/jre-11"
            }
          }
        ]
      ],
      "Properties": {}
    }
  ]

```

## Die JVM-Einstellung mit Apache Hadoop und Hive überschreiben

Im folgenden Beispiel wird gezeigt, wie Sie die JVM für Hadoop und Hive auf Version 17 setzen.

```

[
  {
    "Classification": "hadoop-env",

```

```

    "Configurations": [
      {
        "Classification": "export",
        "Configurations": [],
        "Properties": {
          "JAVA_HOME": "/usr/lib/jvm/jre-17"
        }
      }
    ],
    "Properties": {}
  }
]

```

## Service-Ports

Die folgenden Ports sind YARN- und HDFS-Service-Ports. Diese Einstellungen stellen die Hadoop-Standardwerte dar. Andere Anwendungsservices werden bei Standard-Ports gehostet, sofern nicht anders dokumentiert. Weitere Informationen enthält die Projektdokumentation der Anwendung.

### Port-Einstellungen für YARN und HDFS

Einstellung	Hostname/Port
<code>fs.default.name</code>	Standard () <code>hdfs://<i>emrDeterminedIP</i>:8020</code>
<code>dfs.datanode.address</code>	Standard (0.0.0.0:50010 )
<code>dfs.datanode.http.address</code>	Standard (0.0.0.0:50075 )
<code>dfs.datanode.https.address</code>	Standard (0.0.0.0:50475 )
<code>dfs.datanode.ipc.address</code>	Standard (0.0.0.0:50020 )
<code>dfs.http.address</code>	Standard (0.0.0.0:50070 )
<code>dfs.https.address</code>	Standard (0.0.0.0:50470 )
<code>dfs.secondary.http.address</code>	Standard (0.0.0.0:50090 )
<code>yarn.nodemanager.address</code>	Standard ( <code>\${yarn.nodemanager.hostname}:0</code> )

Einstellung	Hostname/Port
<code>yarn.nodemanager.localizer.address</code>	Standard ( <code>\${yarn.nodemanager.hostname}:8040</code> )
<code>yarn.nodemanager.webapp.address</code>	Standard ( <code>\${yarn.nodemanager.hostname}:8042</code> )
<code>yarn.resourcemanager.address</code>	Standard ( <code>\${yarn.resourcemanager.hostname}:8032</code> )
<code>yarn.resourcemanager.admin.address</code>	Standard ( <code>\${yarn.resourcemanager.hostname}:8033</code> )
<code>yarn.resourcemanager.resource-tracker.address</code>	Standard ( <code>\${yarn.resourcemanager.hostname}:8031</code> )
<code>yarn.resourcemanager.scheduler.address</code>	Standard ( <code>\${yarn.resourcemanager.hostname}:8030</code> )
<code>yarn.resourcemanager.webapp.address</code>	Standard ( <code>\${yarn.resourcemanager.hostname}:8088</code> )
<code>yarn.web-proxy.address</code>	Standard (no-value)
<code>yarn.resourcemanager.hostname</code>	<i>emrDeterminedIP</i>

### Note

Bei dem Begriff *emrDeterminedIP* handelt es sich um eine IP-Adresse, die von der Amazon-EMR-Steuerebene generiert wird. In der neueren Version wurde diese Konvention entfernt, mit Ausnahme der Einstellungen `yarn.resourcemanager.hostname` und `fs.default.name`.

## Anwendungsbenutzer

Die Anwendungen führen Prozesse als deren eigene Benutzer aus. Hive-JVMs werden beispielsweise als Benutzer ausgeführt, Hive, MapReduce JVMs werden als ausgeführt usw. mapred Dies wird im folgenden Prozessstatusbeispiel gezeigt.

```

USER      PID %CPU %MEM    VSZ   RSS TTY      STAT START   TIME COMMAND
hive      6452  0.2  0.7 853684 218520 ?        S1   16:32   0:13 /usr/lib/jvm/
java-openjdk/bin/java -Xmx256m -Dhive.log.dir=/var/log/hive -Dhive.log.file=hive-
metastore.log -Dhive.log.threshold=INFO -Dhadoop.log.dir=/usr/lib/hadoop
hive      6557  0.2  0.6 849508 202396 ?        S1   16:32   0:09 /usr/lib/jvm/java-
openjdk/bin/java -Xmx256m -Dhive.log.dir=/var/log/hive -Dhive.log.file=hive-server2.log
-Dhive.log.threshold=INFO -Dhadoop.log.dir=/usr/lib/hadoop/l
hbase     6716  0.1  1.0 1755516 336600 ?        S1   Jun21   2:20 /usr/lib/jvm/java-
openjdk/bin/java -Dproc_master -XX:OnOutOfMemoryError=kill -9 %p -Xmx1024m -ea -XX:
+UseConcMarkSweepGC -XX:+CMSIncrementalMode -Dhbase.log.dir=/var/
hbase     6871  0.0  0.7 1672196 237648 ?        S1   Jun21   0:46 /usr/lib/jvm/java-
openjdk/bin/java -Dproc_thrift -XX:OnOutOfMemoryError=kill -9 %p -Xmx1024m -ea -XX:
+UseConcMarkSweepGC -XX:+CMSIncrementalMode -Dhbase.log.dir=/var/
hdfs     7491  0.4  1.0 1719476 309820 ?        S1   16:32   0:22 /usr/lib/jvm/java-
openjdk/bin/java -Dproc_namenode -Xmx1000m -Dhadoop.log.dir=/var/log/hadoop-hdfs -
Dhadoop.log.file=hadoop-hdfs-namenode-ip-10-71-203-213.log -Dhadoo
yarn     8524  0.1  0.6 1626164 211300 ?        S1   16:33   0:05 /usr/lib/jvm/java-
openjdk/bin/java -Dproc_proxyserver -Xmx1000m -Dhadoop.log.dir=/var/log/hadoop-yarn -
Dyarn.log.dir=/var/log/hadoop-yarn -Dhadoop.log.file=yarn-yarn-
yarn     8646  1.0  1.2 1876916 385308 ?        S1   16:33   0:46 /usr/lib/jvm/java-
openjdk/bin/java -Dproc_resource manager -Xmx1000m -Dhadoop.log.dir=/var/log/hadoop-yarn
-Dyarn.log.dir=/var/log/hadoop-yarn -Dhadoop.log.file=yarn-y
mapred   9265  0.2  0.8 1666628 260484 ?        S1   16:33   0:12 /usr/lib/jvm/java-
openjdk/bin/java -Dproc_historyserver -Xmx1000m -Dhadoop.log.dir=/usr/lib/hadoop/logs -
Dhadoop.log.file=hadoop.log -Dhadoop.home.dir=/usr/lib/hadoop

```

# Überprüfen von Abhängigkeiten mithilfe des Amazon-EMR-Artefakt-Repositorys

Sie können das Amazon-EMR-Artefakt-Repository verwenden, um Apache-Hive- und Apache-Hadoop-Jobcode anhand der genauen Versionen von Bibliotheken und Abhängigkeiten zu erstellen, die mit bestimmten Amazon-EMR-Release-Versionen verfügbar sind, beginnend mit Amazon-EMR-Version 5.18.0. Die Erstellung auf der Grundlage von Amazon-EMR-Artefakten im Repository hilft dabei, Laufzeit-Klassen-Probleme zu vermeiden, da sichergestellt wird, dass die Versionen der Bibliotheken, auf deren Grundlage der Job erstellt wird, exakt den Versionen entsprechen, die zur Laufzeit auf dem Cluster bereitgestellt werden. Derzeit sind Amazon-EMR-Artefakte nur für Maven-Builds verfügbar.

Fügen Sie zum Zugriff auf das Artefakt-Repository Ihrer Maven-Einstellungsdatei oder der `pom.xml`-Konfigurationsdatei eines bestimmten Projekts die Repository-URL hinzu. Anschließend können Sie die Abhängigkeiten in Ihrer Projektkonfiguration angeben. Verwenden Sie für Abhängigkeitsversionen die unter Komponentenversionen für die gewünschte Version auf [Amazon-EMR-5.x-Versionen](#) aufgeführte Version. Beispielsweise sind Komponentenversionen für die aktuelle Amazon-EMR-Version unter [the section called "5.36.2 Komponentenversionen"](#) verfügbar. Wenn ein Artefakt für Ihr Projekt nicht unter Komponentenversionen aufgeführt ist, geben Sie die Version an, die in diesem Release für Hive und Hadoop aufgeführt ist. Beispielsweise ist für Hadoop-Komponenten in Amazon-EMR-Version 5.18.0 die Version `2.8.4-amzn-1`.

Die Artefakt-Repository-URL hat die folgende Syntax:

```
https://s3-endpoint/region-ID-emr-artifacts/emr-release-label/repos/maven/
```

- *s3-endpoint* ist der Amazon Simple Storage Service (Amazon S3)-Endpunkt der Region für das Repository, und *region-ID* ist die entsprechende Region. Beispiel: `s3.us-west-1.amazonaws.com` und `us-west-1`. Weitere Informationen finden Sie unter Endpunkte für Amazon S3 in Allgemeine Amazon Web Services-Referenz. Es gibt keinen Unterschied hinsichtlich der Artefakte zwischen Regionen. So können Sie die am besten geeignete Region für Ihre Entwicklungsumgebung angeben.
- *emr-release-label* ist das Release-Label für den Amazon EMR-Cluster, auf dem Ihr Code ausgeführt wird. Versionsbezeichnungen haben das Format `emr-x.x.x`, z. B. `emr-5.36.2`. Eine EMR-Veröffentlichungsserie kann mehrere Veröffentlichungen enthalten. Wenn Sie beispielsweise

EMR-Version 5.24.1 verwenden, verwenden Sie die erste EMR-Versionsbezeichnung innerhalb der 5.24-Serie, `emr-5.24.0`, in der Artefakt-Repository-URL:

```
https://s3-endpoint/region-ID-emr-artifacts/emr-5.24.0/repos/maven/
```

### Example Konfiguration für Maven pom.xml

Das pom.xml-Beispiel unten konfiguriert ein Maven-Projekt zum Erstellen auf der Grundlage der `emr-5.18.0` Apache Hadoop- und Apache Hive-Artefakte unter Verwendung des Artefakt-Repositorys in `us-west-1`. Snapshot-Versionen sind im Artefakt-Repository nicht verfügbar, so dass Snapshots im pom.xml deaktiviert sind. Auslassungspunkte (`...`) im folgenden Beispiel zeigen die Auslassung anderer Konfigurationsparameter an. Kopieren Sie diese nicht in Ihr Maven-Projekt.

```
<project>
  ...
  <repositories>
    ...
    <repository>
      <id>emr-5.18.0-artifacts</id>
      <name>EMR 5.18.0 Releases Repository</name>
      <releases>
        <enabled>true</enabled>
      </releases>
      <snapshots>
        <enabled>false</enabled>
      </snapshots>
      <url>https://s3.us-west-1.amazonaws.com/us-west-1-emr-artifacts/emr-5.18.0/repos/
maven/</url>
    </repository>
    ...
  </repositories>
  ...
  <dependencies>
    ...
    <dependency>
      <groupId>org.apache.hive</groupId>
      <artifactId>hive-exec</artifactId>
      <version>2.3.3-amzn-2</version>
    </dependency>
    <dependency>
      <groupId>org.apache.hadoop</groupId>
```

```
<artifactId>hadoop-common</artifactId>  
<version>2.8.4-amzn-1</version>  
</dependency>  
...  
</dependencies>  
  
</project>
```



# EMR File System (EMRFS)

Das EMR File System (EMRFS) ist eine HDFS-Implementierung, die von allen Amazon-EMR-Clustern verwendet wird, um reguläre Dateien von Amazon EMR direkt zu Amazon S3 zu lesen und zu schreiben. EMRFS bietet die praktische Möglichkeit der Speicherung von persistenten Daten in Amazon S3 zur Nutzung mit Hadoop. Gleichzeitig stehen Features wie die Datenverschlüsselung zur Verfügung.

Die Datenverschlüsselung ermöglicht Ihnen das Verschlüsseln von Objekten, die EMRFS zu Amazon S3 schreibt, und ermöglicht EMRFS das Arbeiten mit verschlüsselten Objekten in Amazon S3. Wenn Sie die Amazon-EMR-Version 4.8.0 oder höher verwenden, können Sie Sicherheitskonfigurationen für die Einrichtung der Verschlüsselung von EMRFS-Objekten in Amazon S3 verwenden, zusammen mit anderen Verschlüsselungseinstellungen. Weitere Informationen finden Sie unter [Verschlüsselungsoptionen](#). Wenn Sie eine ältere Version von Amazon EMR verwenden, können Sie Verschlüsselungseinstellungen manuell konfigurieren. Weitere Informationen finden Sie unter [Angeben der Amazon-S3-Verschlüsselung mit EMRFS-Eigenschaften](#).

Amazon S3 bietet eine hohe Konsistenz beim Lesen nach dem Schreiben für alle GET-, PUT- und LIST-Operationen AWS-Regionen. Das bedeutet, dass das, was Sie mit EMRFS schreiben, das ist, was Sie aus Amazon S3 lesen, ohne dass die Leistung beeinträchtigt wird. Weitere Informationen finden Sie unter [Amazon-S3-Datenkonsistenzmodell](#).

Wenn Sie die Amazon-EMR-Version 5.10.0 oder höher verwenden, können Sie verschiedene IAM-Rollen für EMRFS-Anfragen an Amazon S3 verwenden, basierend auf Cluster-Benutzern oder Gruppen oder auf dem Speicherort von EMRFS-Daten in Amazon S3. Weitere Informationen finden Sie unter [Konfigurieren von IAM-Rollen für EMRFS-Anforderungen an Amazon S3](#).

## Warning

Bevor Sie die spekulative Ausführung für Amazon-EMR-Cluster aktivieren, auf denen Apache-Spark-Aufträge ausgeführt werden, lesen Sie bitte die folgenden Informationen. EMRFS umfasst den EMRFS S3-optimierten Committer, eine OutputCommitter Implementierung, die für das Schreiben von Dateien in Amazon S3 bei Verwendung von EMRFS optimiert ist. Wenn Sie das spekulative Ausführungsfeature von Apache Spark mit Anwendungen aktivieren, die Daten in Amazon S3 schreiben und den EMRFS-S3-optimierten Committer nicht verwenden, können Probleme mit der Datenkorrektheit auftreten, die in [SPARK-10063](#) beschrieben sind. Dies kann auftreten, wenn Sie Amazon-EMR-Versionen vor Amazon-EMR-Version 5.19 verwenden oder wenn Sie Dateien mit Formaten wie ORC

und CSV auf Amazon S3 schreiben. Diese Formate werden vom EMRFS-S3-optimierten Committer nicht unterstützt. Eine vollständige Liste der Anforderungen für die Verwendung des EMRFS-S3-optimierten Committers finden Sie unter [Anforderungen für den EMRFS-S3-optimierten Committer](#).

EMRFS Direct Write wird in der Regel verwendet, wenn der EMRFS-S3-optimierte Committer nicht unterstützt wird, z. B. wenn Folgendes geschrieben wird:

- Ein anderes Ausgabeformat als Parquet, z. B. ORC oder Text.
- Hadoop-Dateien, die die Spark-RDD-API verwenden.
- Parquet mit Hive. Siehe unter [Konvertierung der Parquet-Tabelle in Hive Metastore](#).

EMRFS-Direct-Write-Verfahren wird in den folgenden Szenarien nicht verwendet:

- Wenn der S3-optimierte EMRFS-Committer aktiviert ist. Sehen Sie unter [Anforderungen für den S3-optimierten EMRFS-Committer](#).
- Beim Schreiben dynamischer Partitionen, bei denen die `partitionOverwriteMode` Einstellung auf dynamisch gesetzt ist.
- Beim Schreiben in benutzerdefinierte Partitionsspeicherorte, z. B. an Speicherorte, die nicht der Hive-Standardkonvention für Partitionen entsprechen.
- Wenn Sie andere Dateisysteme als EMRFS verwenden, z. B. beim Schreiben in HDFS oder beim Verwenden des S3A-Dateisystems.

Um festzustellen, ob Ihre Anwendung Direct Write in Amazon EMR 5.14.0 oder höher verwendet, aktivieren Sie die Spark-INFO-Protokollierung. Wenn eine Protokollzeile mit dem Text „Direct Write: ENABLED“ entweder in den Spark-Treiberprotokollen oder in den Spark-Executor-Container-Protokollen vorhanden ist, dann hat Ihre Spark-Anwendung mit Direct Write geschrieben.

Standardmäßig ist die spekulative Ausführung auf Amazon EMRClusters OFF aktiviert. Wir empfehlen dringend, die spekulative Ausführung nicht zu aktivieren, wenn diese beiden Bedingungen zutreffen:

- Sie schreiben Daten in Amazon S3.
- Daten werden in einem anderen Format als Apache Parquet oder im Apache-Parquet-Format ohne den für EMRFS-S3-optimierten Committer geschrieben.

Wenn Sie die spekulative Ausführung von Spark aktivieren und Daten mithilfe von EMRFS Direct Write auf Amazon S3 schreiben, kann es zu zeitweiligem Datenverlust kommen. Wenn Sie Daten in HDFS schreiben oder Daten mit dem EMRFS-S3-optimierten Committer in Parquet schreiben, verwendet Amazon EMR kein direktes Schreiben und dieses Problem tritt nicht auf.

Wenn Sie Daten in Formaten schreiben müssen, die EMRFS Direct Write von Spark nach Amazon S3 verwenden und spekulative Ausführung verwenden, empfehlen wir, in HDFS zu schreiben und dann die Ausgabedateien mit S3DistCP nach Amazon S3 zu übertragen.

## Themen

- [Consistent view](#)
- [Zugriff auf EMRFS-Daten in Amazon S3 genehmigen](#)
- [Verwaltung des Standardendpunkts AWS Security Token Service](#)
- [Angaben der Amazon-S3-Verschlüsselung mit EMRFS-Eigenschaften](#)

## Consistent view

### Warning

Am 1. Juni 2023 wird die Standardunterstützung für zukünftige Amazon-EMR-Versionen für EMRFS konsistente Ansicht eingestellt. Die EMRFS-konsistente Ansicht wird auch weiterhin für bestehende Versionen funktionieren.

Mit der Veröffentlichung von Amazon S3 Strong read-after-write Consistency am 1. Dezember 2020 müssen Sie EMRFS Consistent View (EMRFS CV) nicht mehr für Ihre Amazon EMR-Cluster verwenden. EMRFS CV ist eine optionale Funktion, mit der Amazon EMR-Cluster Amazon S3 S3-Objekte auflisten und read-after-write Konsistenz überprüfen können. Wenn Sie einen Cluster erstellen und EMRFS CV aktiviert ist, erstellt Amazon EMR eine Amazon DynamoDB-Datenbank zum Speichern von Objektmetadaten, die zur Nachverfolgung der Liste und read-after-write Konsistenz von S3-Objekten verwendet werden. Sie können jetzt EMRFS CV deaktivieren und die DynamoDB-Datenbank löschen, sodass Ihnen keine zusätzlichen Kosten entstehen. In den

folgenden Verfahren wird erklärt, wie Sie das CV-Feature prüfen, ausschalten und die DynamoDB-Datenbank löschen, die das Feature verwendet.

Um zu überprüfen, ob Sie das EMRFS-CV-Feature verwenden

1. Navigieren Sie zur Registerkarte Konfiguration. Wenn Ihr Cluster die folgende Konfiguration hat, verwendet er EMRFS CV.

```
Classification=emrfs-site,Property=fs.s3.consistent,Value=true
```

2. [Verwenden Sie alternativ den, um Ihren Cluster AWS CLI mit der API zu beschreiben. describe-cluster](#) Wenn die Ausgabe `fs.s3.consistent: true` enthält, verwendet Ihr Cluster EMRFS CV.

So deaktivieren Sie EMRFS CV auf Ihren Amazon-EMR-Clustern

Verwenden Sie eine der folgenden drei Optionen, um das EMRFS-CV-Feature zu deaktivieren. Sie sollten diese Optionen in Ihrer Testumgebung testen, bevor Sie sie auf Ihre Produktionsumgebungen anwenden.

1. Um Ihren vorhandenen Cluster zu beenden und einen neuen Cluster ohne EMRFS-CV-Optionen zu starten.
  - a. Bevor Sie Ihren Cluster beenden, stellen Sie sicher, dass Sie Ihre Daten sichern und Ihre Benutzer benachrichtigen.
  - b. Befolgen Sie die Anweisungen unter [Einen Cluster beenden](#), um den Cluster zu beenden.
  - c. Wenn Sie die Amazon-EMR-Konsole verwenden, um einen neuen Cluster zu erstellen, navigieren Sie zu Erweiterte Optionen. Deaktivieren Sie im Abschnitt Softwareeinstellungen bearbeiten die Option zum Aktivieren von EMRFS CV. Wenn das Kontrollkästchen für die konsistente EMRFS-Ansicht verfügbar ist, lassen Sie es deaktiviert.
  - d. Wenn Sie AWS CLI früher einen neuen Cluster mit der [create-clusterAPI](#) erstellen, verwenden Sie nicht die `--emrfs` Option, die EMRFS CV aktiviert.
  - e. Wenn Sie ein SDK verwenden oder einen neuen Cluster erstellen AWS CloudFormation möchten, verwenden Sie keine der Konfigurationen, die unter [Konsistente Ansicht konfigurieren](#) aufgeführt sind.
2. So klonen Sie einen Cluster und entfernen EMRFS CV
  - a. Klicken Sie in der Amazon-EMR-Konsole auf den Cluster, der EMRFS-CV verwendet.

- b. Wählen Sie oben auf der Seite Cluster Details (Clusterdetails) auf Clone (Klonen).
- c. Wählen Sie Zurück und navigieren Sie zu Schritt 1: Software und Schritte.
- d. Entfernen Sie unter Softwareeinstellungen bearbeiten die Datei EMRFS CV. Löschen Sie unter Konfiguration bearbeiten die folgenden Konfigurationen in der `emrfs-site`-Klassifizierung. Wenn Sie JSON aus einem S3-Bucket laden, müssen Sie Ihr S3-Objekt ändern.

```
[
  {
    "classification":
      "emrfs-site",
    "properties": {
      "fs.s3.consistent.retryPeriodSeconds": "10",
      "fs.s3.consistent": "true",
      "fs.s3.consistent.retryCount": "5",
      "fs.s3.consistent.metadata.tableName": "EmrFSMetadata"
    }
  }
]
```

3. Um EMRFS CV aus einem Cluster zu entfernen, der Instance-Gruppen verwendet
  - a. Verwenden Sie den folgenden Befehl, um zu überprüfen, ob ein einzelner EMR-Cluster die DynamoDB-Tabelle verwendet, die mit EMRFS CV verknüpft ist, oder ob mehrere Cluster die Tabelle gemeinsam nutzen. Der Tabellename ist in `fs.s3.consistent.metadata.tableName` angegeben, wie unter [Konsistente Ansicht konfigurieren beschrieben](#). Der von EMRFS CV verwendete Standardtabellename ist `EmrFSMetadata`.

```
aws emr describe-cluster --cluster-id j-XXXXX | grep
fs.s3.consistent.metadata.tableName
```

- b. Wenn Ihr Cluster Ihre DynamoDB-Datenbank nicht mit einem anderen Cluster teilt, verwenden Sie den folgenden Befehl, um den Cluster neu zu konfigurieren und EMRFS CV zu deaktivieren. Weitere Informationen finden Sie unter [Neu-Konfiguration für eine Instance-Gruppe in einem aktiven Cluster](#).

```
aws emr modify-instance-groups --cli-input-json file://disable-emrfs-1.json
```

Mit diesem Befehl wird die Datei geöffnet, die Sie ändern möchten. Ändern Sie die Datei mit den folgenden Konfigurationen.

```
{
  "ClusterId": "j-xxxx",
  "InstanceGroups": [
    {
      "InstanceGroupId": "ig-xxxx",
      "Configurations": [
        {
          "Classification": "emrfs-site",
          "Properties": {
            "fs.s3.consistent": "false"
          }
        },
        "Configurations": []
      ]
    }
  ]
}
```

- c. Wenn Ihr Cluster die DynamoDB-Tabelle mit einem anderen Cluster gemeinsam nutzt, deaktivieren Sie EMRFS CV auf allen Clustern zu einem Zeitpunkt, zu dem keine Cluster Objekte am gemeinsam genutzten S3-Standort ändern.

Löschen Sie Amazon-DynamoDB-Ressourcen, die mit EMRFS CV verknüpft sind, wie folgt

Nachdem Sie EMRFS CV aus Ihren Amazon-EMR-Clustern entfernt haben, löschen Sie die DynamoDB-Ressourcen, die mit EMRFS CV verknüpft sind. Solange Sie dies nicht tun, fallen für Sie weiterhin DynamoDB-Gebühren im Zusammenhang mit EMRFS CV an.

1. Überprüfen Sie die CloudWatch Metriken für Ihre DynamoDB-Tabelle und stellen Sie sicher, dass die Tabelle von keinem Cluster verwendet wird.
2. Löschen Sie die DynamoDB-Tabelle.

```
aws dynamodb delete-table --table-name <your-table-name>
```

Löschen Sie Amazon-SQS-Ressourcen, die mit EMRFS CV verknüpft sind, wie folgt

1. Wenn Sie Ihren Cluster so konfiguriert haben, dass Inkonsistenzbenachrichtigungen an Amazon SQS gesendet werden, können Sie alle SQS-Warteschlangen löschen.
2. Suchen Sie den in `fs.s3.consistent.notification.SQS.queueName` angegebenen Amazon-SQS-Warteschlangennamen, wie unter [Konsistente Ansicht konfigurieren](#) beschrieben. Das Standardformat für Warteschlangennamen ist `EMRFS-Inconsistency-<j-cluster ID>`.

```
aws sqs list-queues | grep 'EMRFS-Inconsistency'  
aws sqs delete-queue --queue-url <your-queue-url>
```

So beenden Sie die Verwendung der EMRFS-CLI

- Die [EMRFS-CLI verwaltet die Metadaten, die EMRFS CV](#) generiert. Da die Standardunterstützung für EMRFS CV in zukünftigen Versionen von Amazon EMR ausläuft, wird auch die Unterstützung für die EMRFS-CLI eingestellt.

Themen

- [Die konsistente Ansicht aktivieren](#)
- [Verstehen, wie mit der konsistenten Ansicht von EMRFS Objekte in Amazon S3 nachverfolgt werden](#)
- [Logik für Wiederholversuche](#)
- [Konsistente Ansicht der Metadaten in EMRFS](#)
- [Konsistenzbenachrichtigungen für CloudWatch und Amazon SQS konfigurieren](#)
- [Konfigurieren der konsistenten Ansicht](#)
- [Die EMRFS-CLI-Befehlsreferenz](#)

## Die konsistente Ansicht aktivieren

Sie können die serverseitige Amazon S3 S3-Verschlüsselung oder die konsistente Ansicht für EMRFS mithilfe der AWS Management Console Konfigurationsklassifizierung AWS CLI, oder aktivieren. `emrfs-site`

## So konfigurieren Sie die konsistente Ansicht mit der Konsole

1. Navigieren Sie zur neuen Amazon-EMR-Konsole und wählen Sie in der Seitennavigation die Option Zur alten Konsole wechseln aus. Weitere Informationen darüber, was Sie erwartet, wenn Sie zur alten Konsole wechseln, finden Sie unter [Verwenden der alten Konsole](#).
2. Wählen Sie Create Cluster (Cluster erstellen) und Go to advanced options (Zu erweiterten Optionen) aus.
3. Wählen Sie in Step 1: Software and Steps (Schritt 1: Software und Schritte) und Step 2: Hardware (Schritt 2: Hardware) Einstellungen aus.
4. Wählen Sie in Step 3: General Cluster Settings (Schritt 3: Cluster-Standard-einstellungen) in Additional Options (Zusätzliche Optionen) die Option EMRFS consistent view (EMRFS-konsistente Ansicht) aus.
5. Geben Sie in EMRFS Metadata store (EMRFS-Metadaten-speicher) den Namen Ihres Metadaten-speichers ein. Der Standardwert ist **EmrFSMetadata**. Wenn die Tabelle EmrFSMetadata nicht vorhanden ist, wird sie in DynamoDB für Sie erstellt.

### Note

Amazon EMR entfernt die EMRFS-Metadaten nicht automatisch aus DynamoDB, wenn der Cluster beendet wird.

6. Geben Sie in Number of retries (Anzahl der Wiederholungen) einen Ganzzahlwert ein. Wenn eine Inkonsistenz festgestellt wird, versucht EMRFS, Amazon S3 so oft wie möglich aufzurufen. Der Standardwert ist **5**.
7. Geben Sie in Retry period (in seconds) (Wiederholungszeitraum (in Sekunden)) einen Ganzzahlwert ein. Dies ist die Zeitspanne, die EMRFS zwischen Wiederholungsversuchen wartet. Der Standardwert ist **10**.

### Note

Bei nachfolgenden Wiederholversuchen wird ein exponentielles Backoff verwendet.

Um einen Cluster mit aktivierter konsistenter Ansicht zu starten, verwenden Sie AWS CLI

Wir empfehlen Ihnen, die aktuelle Version von zu installieren AWS CLI. Die neueste Version können Sie unter <https://aws.amazon.com/cli/> herunterladen.



• **Note**

Linux-Zeilenumbruchzeichen (\) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (^).

```
aws emr create-cluster --instance-type m5.xlarge --instance-count 3 --emrfs  
  Consistent=true \  
  --release-label emr-7.1.0 --ec2-attributes KeyName=myKey
```

Um zu überprüfen, ob die konsistente Ansicht aktiviert ist, verwenden Sie AWS Management Console

- Um in der Konsole zu überprüfen, ob die konsistente Ansicht aktiviert ist, navigieren Sie zu Cluster List (Cluster-Liste) und wählen den Namen Ihres Clusters aus, um die Cluster Details (Cluster-Details) anzuzeigen. Das Feld "EMRFS consistent view" hat den Wert Enabled oder Disabled.

So prüfen Sie durch Untersuchen der Datei **emrfs-site.xml**, ob die konsistente Ansicht aktiviert ist

- Sie können überprüfen, ob die Konsistenz aktiviert ist, indem Sie die Konfigurationsdatei `emrfs-site.xml` auf dem Master-Knoten des Clusters prüfen. Wenn der boolesche Wert für `fs.s3.consistent` auf `true` festgelegt ist, ist die konsistente Ansicht für Dateisystem-Operationen, an denen Amazon S3 beteiligt ist, aktiviert.

## Verstehen, wie mit der konsistenten Ansicht von EMRFS Objekte in Amazon S3 nachverfolgt werden

EMRFS erstellt eine konsistente Ansicht der Objekte in Amazon S3, indem es Informationen über diese Objekte zu den EMRFS-Metadaten hinzufügt. EMRFS fügt diese Listen seinen Metadaten hinzu, wenn Folgendes zutrifft:

- Ein Objekt wird von EMRFS im Laufe eines Amazon-EMR-Auftrags geschrieben.
- Ein Objekt wird unter Verwendung der EMRFS-CLI mit EMRFS-Metadaten synchronisiert bzw. in die EMRFS-Metadaten importiert.

Objekte, die von EMRFS gelesen werden, werden nicht automatisch zu den Metadaten hinzugefügt. Wenn EMRFS ein Objekt löscht, verbleibt eine Liste in den Metadaten mit dem Status „Gelöscht“, bis diese Auflistung über die EMRFS-CLI bereinigt wird. Weitere Informationen zur CLI finden Sie unter [Die EMRFS-CLI-Befehlsreferenz](#). Weitere Informationen zum Bereinigen von Listen in den EMRFS-Metadaten finden Sie unter [Konsistente Ansicht der Metadaten in EMRFS](#).

Für jede Amazon-S3-Operation überprüft EMRFS die Metadaten auf Informationen über die Objekte in der konsistenten Ansicht. Wenn EMRFS feststellt, dass Amazon S3 während einer dieser Operationen inkonsistent ist, wird die Operation entsprechend den in den `emrfs-site`-Konfigurationseigenschaften definierten Parametern wiederholt. Nachdem EMRFS alle Wiederholversuche ausgeschöpft hat, wirft es entweder eine `ConsistencyException` auf, oder es protokolliert die Ausnahme und setzt den Workflow fort. Weitere Informationen zur Wiederholungslogik finden Sie unter [Logik für Wiederholversuche](#). Sie finden die `ConsistencyExceptions` in Ihren Protokollen, zum Beispiel:

- `listStatus`: Kein Amazon S3 Objekt für Metadatenelement `/S3_bucket/dir/object`
- `getFileStatus`: Der Schlüssel `dir/file` ist in den Metadaten vorhanden, aber nicht in Amazon S3

Wenn Sie ein mit der konsistenten Ansicht von EMRFS nachverfolgtes Objekt direkt in Amazon S3 löschen, behandelt EMRFS dieses Objekt als inkonsistent, da es weiterhin in den Metadaten als in Amazon S3 vorhanden angezeigt wird. Wenn Ihre Metadaten nicht mehr mit den Objekten synchronisiert sind, die EMRFS in Amazon S3 nachverfolgt, können Sie den Unterbefehl `sync` der EMRFS-CLI ausführen, um die Metadaten zurückzusetzen, sodass sie den aktuellen Stand in Amazon S3 reflektieren. Um Abweichungen zwischen Metadaten und Amazon S3 zu entdecken, verwenden Sie den Befehl `diff`. EMRFS verfügt nur über eine konsistente Ansicht der Objekte, auf die in den Metadaten verwiesen wird; im gleichen Amazon-S3-Pfad können andere Objekte vorhanden sein, die nicht verfolgt werden. Wenn EMRFS die Objekte in einem Amazon-S3-Pfad auflistet, wird die Gesamtmenge der Objekte zurückgegeben, die in den Metadaten nachverfolgt werden und im Amazon-S3-Pfad vorhanden sind.

## Logik für Wiederholversuche

EMRFS versucht mit einer bestimmten Anzahl an Wiederholversuchen, die Listenkonsistenz für Objekte zu überprüfen, die in den Metadaten verfolgt werden. Der Standardwert ist 5. Wenn die Anzahl der Wiederholversuche überschritten wird, gibt der ursprüngliche Auftrag einen Fehler zurück, es sei denn, `fs.s3.consistent.throwExceptionOnInconsistency` ist auf `false` festgelegt, sodass die Objekte lediglich protokolliert werden, die als inkonsistent

verfolgt wurden. EMRFS verwendet standardmäßig eine Wiederholungsrichtlinie mit exponentiellem Backoff, Sie können jedoch auch eine feste Richtlinie verwenden. Möglicherweise möchten die Benutzer ihre Versuche für einen bestimmten Zeitraum wiederholen, bevor sie den restlichen Auftrag ausführen, ohne dass eine Ausnahme ausgelöst wird. Sie können dies erreichen, indem sie `fs.s3.consistent.throwExceptionOnInconsistency` auf `false`, `fs.s3.consistent.retryPolicyType` auf `fixed` und `fs.s3.consistent.retryPeriodSeconds` auf den gewünschten Wert festlegen. Im folgenden Beispiel wird ein Cluster mit aktivierter Konsistenz erstellt. Dabei werden Inkonsistenzen protokolliert und es wird ein festes Wiederholungsintervall von 10 Sekunden festgelegt:

### Example Festlegen des Wiederholungszeitraums auf einen festen Wert

```
aws emr create-cluster --release-label emr-7.1.0 \  
--instance-type m5.xlarge --instance-count 1 \  
--emrfs Consistent=true,Args=[fs.s3.consistent.throwExceptionOnInconsistency=false,  
fs.s3.consistent.retryPolicyType=fixed,fs.s3.consistent.retryPeriodSeconds=10] --ec2-  
attributes KeyName=myKey
```

#### Note

Linux-Zeilenfortsetzungszeichen (\) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (^).

Weitere Informationen finden Sie unter [Consistent view](#).

## EMRFS-Konfigurationen für IMDS rufen Regionen ab

EMRFS nutzt den IMDS (Instance-Metadaten-Service), um die Instance-Region und Amazon S3, DynamoDB oder Endpunkte abzurufen. AWS KMS IMDS hat jedoch ein Limit für die Anzahl der Anfragen, die es verarbeiten kann, und Anfragen, die dieses Limit überschreiten, schlagen fehl. Dieses IMDS-Limit kann dazu führen, dass EMRFS-Initialisierungen fehlschlagen und die Abfrage oder der Befehl fehlschlagen. Sie können den folgenden Mechanismus für zufällige exponentielle Backoff-Wiederholungen und die Konfigurationseigenschaften einer Fallback-Region in `emrfs-site.xml` verwenden, um das Szenario zu beheben, in dem alle Wiederholungsversuche fehlschlagen.

```
<property>
```

```
<name>fs.s3.region.retryCount</name>
<value>3</value>
<description>
Maximum retries that would be attempted to get AWS region.
</description>
</property>
<property>
  <name>fs.s3.region.retryPeriodSeconds</name>
  <value>3</value>
  <description>
Base sleep time in second for each get-region retry.
  </description>
</property>
<property>
  <name>fs.s3.region.fallback</name>
  <value>us-east-1</value>
  <description>
Fallback to this region after maximum retries for getting AWS region have been
reached.
  </description>
</property>
```

## Konsistente Ansicht der Metadaten in EMRFS

Die konsistente Ansicht von EMRFS verfolgt die Konsistenz unter Verwendung einer DynamoDB-Tabelle nach, um Objekte in Amazon S3 nachzuverfolgen, die mit EMRFS synchronisiert oder erstellt wurden. Mithilfe dieser Metadaten werden alle Operationen (Lesen, Schreiben, Aktualisieren und Kopieren) verfolgt. Sie enthalten keine tatsächlichen Inhalte. Anhand der Metadaten wird überprüft, ob die Objekte oder Metadaten, die von Amazon S3 eingehen, den erwarteten Werten entsprechen. Diese Bestätigung gibt EMRFS die Möglichkeit, die Listenkonsistenz und read-after-write Konsistenz für neue Objekte, die EMRFS in Amazon S3 schreibt, oder für Objekte, die mit EMRFS synchronisiert wurden, zu überprüfen. Mehrere Cluster können dieselben Metadaten verwenden.

So fügen Sie Einträge zu den Metadaten hinzu

Sie können die Unterbefehle `import` oder `sync` verwenden, um Metadaten Einträge hinzuzufügen. `sync` gibt den Status der Amazon-S3-Objekte in einem Pfad an, während `import` ausschließlich zum Hinzufügen neuer Einträge zu den Metadaten verwendet wird. Weitere Informationen finden Sie unter [Die EMRFS-CLI-Befehlsreferenz](#).

So überprüfen Sie Unterschiede zwischen Metadaten und Objekten in Amazon S3

Um eine Prüfung auf Abweichungen zwischen den Metadaten und Amazon S3 auszuführen, verwenden Sie den Unterbefehl `diff` der EMRFS-CLI. Weitere Informationen finden Sie unter [Die EMRFS-CLI-Befehlsreferenz](#).

So erkennen Sie, ob Metadatenoperationen gedrosselt werden

EMRFS definiert die maximale Standarddurchsatzkapazität für die Metadaten für Lese- und Schreiboperationen mit jeweils 500 und 100 Einheiten. Eine große Anzahl von Objekten oder Buckets kann dazu führen, dass Vorgänge dieses Limit überschreiten. Sie werden dann von DynamoDB gedrosselt. Zum Beispiel kann es bei einer Anwendung dazu kommen, dass EMRFS eine `ProvisionedThroughputExceededException` auslöst, wenn Sie eine Operation durchführen, die diese Kapazitätsgrenzen überschreitet. Bei der Drosselung versucht das EMRFS-CLI-Tool, mit [exponentiellem Backoff](#) erneut in die DynamoDB-Tabelle zu schreiben, bis die Operation abgeschlossen ist oder wenn der Höchstwert für die Wiederholversuche für das Schreiben von Objekten von Amazon EMR nach Amazon S3 erreicht ist.

Sie können Ihre eigene Durchsatzkapazitätsgrenzen konfigurieren. DynamoDB hat jedoch strenge Partitions Grenzen von 3 000 Lesekapazitätseinheiten (RCUs) und 1 000 Schreibkapazitätseinheiten (WCUs) pro Sekunde für Lese- und Schreibvorgänge. Um durch Drosselung verursachte sync-Ausfälle zu vermeiden, empfehlen wir, den Durchsatz für Lesevorgänge auf weniger als 3 000 RCUs und Schreibvorgänge auf weniger als 1 000 WCUs zu begrenzen. Anweisungen zur Festlegung benutzerdefinierter Durchsatzkapazitätsgrenzen finden Sie unter [Konfigurieren der konsistenten Ansicht](#).

Sie können auch CloudWatch Amazon-Metriken für Ihre EMRFS-Metadaten in der DynamoDB-Konsole anzeigen, wo Sie die Anzahl der gedrosselten Lese- und Schreibvorgängen sehen können. Wenn für gedrosselte Anforderungen ein Wert ungleich null vorliegt, profitiert Ihre Anwendung möglicherweise von einer Erhöhung der zugewiesenen Durchsatzkapazität für Lese- oder Schreiboperationen. Sie realisieren ggf. auch dann eine höhere Performance, wenn Sie sehen, dass sich Ihre Operationen dem zugewiesenen Höchstwert für die Durchsatzkapazität für Lese- und Schreibvorgänge für einen längeren Zeitraum nähern.

Durchsatzmerkmale für bedeutende EMRFS-Operationen

Die Standardeinstellung für Lese- und Schreibvorgänge beträgt 400 bzw. 100 Durchsatzkapazitätseinheiten. Die folgenden Leistungsmerkmale geben Ihnen eine Vorstellung davon, welcher Durchsatz für bestimmte Operationen erforderlich ist. Diese Tests wurden unter Verwendung eines `m3.large`-Clusters mit einem einzigen Knoten durchgeführt. Bei allen Operationen handelte es sich um Single-Thread-Vorgänge. Die Performance kann je nach

Charakteristika der jeweiligen Anwendung sehr unterschiedlich ausfallen und etwas Experimentieren erfordern, um die Dateisystemoperationen zu optimieren.

Operation	Durchschnittlich read-per-second	Durchschnittlich write-per-second
create (Objekt)	26.79	6.70
delete (Objekt)	10.79	10.79
delete (Verzeichnis mit 1000 Objekten)	21.79	338.40
getFileStatus(Objekt)	34.70	0
getFileStatus(Verzeichnis)	19.96	0
listStatus (Verzeichnis mit 1 Objekt)	43.31	0
listStatus (Verzeichnis mit 10 Objekten)	44.34	0
listStatus (Verzeichnis mit 100 Objekten)	84.44	0
listStatus (Verzeichnis mit 1.000 Objekten)	308.81	0
listStatus (Verzeichnis mit 10.000 Objekten)	416.05	0
listStatus (Verzeichnis mit 100.000 Objekten)	823.56	0
listStatus (Verzeichnis mit 1 Million Objekten)	882.36	0
mkdir (durchgehend für 120 Sekunden)	24.18	4.03

Operation	Durchschnittlich read-per-second	Durchschnittlich write-per-second
mkdir	12.59	0
rename (Objekt)	19.53	4.88
rename (Verzeichnis mit 1000 Objekten)	23.22	339.34

So übermitteln Sie einen Schritt, der alle alten Daten aus Ihren Metadaten entfernt

Unter Umständen möchten Benutzer bestimmte Einträge in den DynamoDB-basierten Metadaten entfernen. So können Sie die Speicherkosten im Zusammenhang mit der Tabelle reduzieren. Die Benutzer haben die Möglichkeit, mit dem Unterbefehl `delete` der EMRFS-Befehlszeilenschnittstelle bestimmte Einträge manuell oder programmgesteuert zu bereinigen. Wenn Sie jedoch Einträge aus den Metadaten löschen, führt EMRFS keine Konsistenzprüfungen mehr durch.

Die programmgesteuerte Bereinigung nach Abschluss eines Auftrags kann durch Übermittlung eines letzten Schritts zu Ihrem Cluster erfolgen, wodurch ein Befehl in der EMRFS-Befehlszeilenschnittstelle ausgeführt wird. Geben Sie z. B. den folgenden Befehl ein, um einen Schritt zum Löschen aller Einträge an Ihren Cluster zu übermitteln, die älter als zwei Tage sind.

```
aws emr add-steps --cluster-id j-2AL4XXXXXX5T9 --steps Name="emrfsCLI",Jar="command-
runner.jar",Args=["emrfs","delete","--time","2","--time-unit","days"]
{
  "StepIds": [
    "s-B12345678902"
  ]
}
```

Verwenden Sie den zurückgegebenen `StepId` Wert, um in den Protokollen nach dem Ergebnis des Vorgangs zu suchen.

## Konsistenzbenachrichtigungen für CloudWatch und Amazon SQS konfigurieren

Sie können CloudWatch Metriken und Amazon SQS SQS-Nachrichten in EMRFS für eventuelle Konsistenzprobleme in Amazon S3 aktivieren.

## CloudWatch

Wenn CloudWatch Metriken aktiviert sind, wird jedes Mal, wenn ein FileSystem API-Aufruf aufgrund der eventuellen Konsistenz von Amazon S3 fehlschlägt, eine Metrik namens Inconsistency übertragen.

So zeigen Sie CloudWatch Metriken für eventuelle Konsistenzprobleme in Amazon S3 an

Um die Inkonsistenz-Metrik in der CloudWatch Konsole anzuzeigen, wählen Sie die EMRFS-Metriken und dann ein Paar mit dem Namen JobFlowId/Metric aus. Beispiel: j-162XXXXXXM2CU ListStatus, j-162XXXXXXM2CU GetFileStatus usw.

1. [Öffnen Sie die CloudWatch Konsole unter https://console.aws.amazon.com/cloudwatch/](https://console.aws.amazon.com/cloudwatch/).
2. Wählen Sie im Dashboard (Dashboard) im Abschnitt Metrics (Metriken) die Option EMRFS (EMRFS) aus.
3. Wählen Sie im Bereich Job Flow Metrics ein oder mehrere JobFlowId/Metric Name-Paare aus. Im Fenster darunter wird eine grafische Darstellung der Metriken angezeigt.

## Amazon SQS

Wenn Amazon-SQS-Benachrichtigungen aktiviert sind, wird bei Initialisierung von EMRFS eine Amazon SQS-Warteschlange namens EMRFS-Inconsistency-<jobFlowId> erstellt. Amazon-SQS-Nachrichten werden per Push-Verfahren in die Warteschlange gestellt, wenn ein FileSystem-API-Aufruf aufgrund eines eventuellen Konsistenzproblems von Amazon S3 fehlschlägt. Die Nachricht enthält Informationen wie API JobFlowId, eine Liste inkonsistenter Pfade, einen Stack-Trace usw. Nachrichten können mit der Amazon-SQS-Konsole oder dem EMRFS-Befehl `read-sqs` gelesen werden.

Verwalten von Amazon-SQS-Nachrichten bei eventuellen Konsistenzproblemen von Amazon S3

Amazon-SQS-Nachrichten bei eventuellen Konsistenzproblemen in Amazon S3 können über die EMRFS-CLI gelesen werden. Um Nachrichten aus einer EMRFS-Amazon-SQS-Warteschlange zu lesen, geben Sie den Befehl `read-sqs` ein und geben Sie für die resultierende Ausgabedatei einen Speicherort im lokalen Dateisystem des Hauptknotens an.

Sie können eine EMRFS-Amazon-SQS-Warteschlange mit dem Befehl `delete-sqs` löschen.

1. Zum Lesen von Nachrichten aus einer Amazon-SQS-Warteschlange geben Sie den folgenden Befehl ein. Ersetzen Sie *Warteschlangename* durch den Namen der Amazon-SQS-



Warteschlange, die Sie konfiguriert haben, und ersetzen Sie */path/filename* durch den Pfad zur Ausgabedatei:

```
emrfs read-sqs --queue-name queuename --output-file /path/filename
```

Um z. B. Amazon-SQS-Nachrichten aus der Standardwarteschlange zu lesen und auszugeben, geben Sie Folgendes ein:

```
emrfs read-sqs --queue-name EMRFS-Inconsistency-j-162XXXXXXM2CU --output-file /path/filename
```

#### Note

Sie können auch die Verknüpfungen `-q` und `-o` anstelle von `--queue-name` und `--output-file` verwenden.

2. Zum Löschen einer Amazon-SQS-Warteschlange geben Sie den folgenden Befehl ein:

```
emrfs delete-sqs --queue-name queuename
```

Um z. B. die Standardwarteschlange zu löschen, geben Sie Folgendes ein:

```
emrfs delete-sqs --queue-name EMRFS-Inconsistency-j-162XXXXXXM2CU
```

#### Note

Sie können auch das Kürzel `-q` anstelle von `--queue-name` verwenden.

## Konfigurieren der konsistenten Ansicht

Sie können zusätzliche Einstellungen für die konsistente Ansicht konfigurieren, indem Sie sie mithilfe von Konfigurationseigenschaften für `emrfs-site`-Eigenschaften bereitstellen. Sie können beispielsweise einen anderen Standard-DynamoDB-Durchsatz wählen, indem Sie die folgenden Argumente an die CLI-`--emrfs`-Option übergeben, die `emrfs-site`-Konfigurationsklassifizierung (nur Amazon EMR Version 4.x und höher) oder eine Bootstrap-Aktion zur Konfiguration der Datei `emrfs-site.xml` auf dem Hauptknoten verwenden:

## Example Ändern der Standardwerte zum Lesen und Schreiben von Metadaten bei Cluster-Start

```
aws emr create-cluster --release-label emr-7.1.0 --instance-type m5.xlarge \  
--emrfs Consistent=true,Args=[fs.s3.consistent.metadata.read.capacity=600,\  
fs.s3.consistent.metadata.write.capacity=300] --ec2-attributes KeyName=myKey
```

Verwenden Sie alternativ die folgende Konfigurationsdatei und speichern Sie sie lokal oder in Amazon S3:

```
[  
  {  
    "Classification": "emrfs-site",  
    "Properties": {  
      "fs.s3.consistent.metadata.read.capacity": "600",  
      "fs.s3.consistent.metadata.write.capacity": "300"  
    }  
  }  
]
```

Verwenden Sie die Konfiguration, die Sie erstellt haben, mit der folgenden Syntax:

```
aws emr create-cluster --release-label emr-7.1.0 --applications Name=Hive \  
--instance-type m5.xlarge --instance-count 2 --configurations file://./myConfig.json
```

### Note

Linux-Zeilenfortsetzungszeichen (\) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (^).

Die folgenden Optionen können mithilfe von Konfigurationen oder AWS CLI `--emrfs` Argumenten festgelegt werden. Informationen über diese Argumente finden Sie in der [AWS CLI -Befehlsreferenz](#).

### **emrfs-site.xml**-Eigenschaften für die konsistente Ansicht

Eigenschaft	Standardwert	Beschreibung
<code>fs.s3.consistent</code>	<b>false</b>	

Eigenschaft	Standardwert	Beschreibung
		Wenn diese Eigenschaft auf <b>true</b> festgelegt wurde, wird EMRFS für die Verwendung von DynamoDB zur Bereitstellung von Konsistenz konfiguriert.
<code>fs.s3.consistent.retryPolicyType</code>	<b>exponential</b>	Diese Eigenschaft gibt die Richtlinie an, die für Wiederholversuche bei Konsistenzproblemen verwendet werden soll. Optionen: exponentiell, fest oder keine.
<code>fs.s3.consistent.retryPeriodSeconds</code>	<b>1</b>	Diese Eigenschaft legt die Wartezeit zwischen den Wiederholversuchen bei Konsistenzproblemen fest.
<code>fs.s3.consistent.retryCount</code>	<b>10</b>	Diese Eigenschaft legt die maximale Anzahl von Wiederholversuchen bei erkannten Inkonsistenzen fest.
<code>fs.s3.consistent.throwExceptionOnInconsistency</code>	<b>true</b>	Diese Eigenschaft bestimmt, ob eine Konsistenzausnahme ausgelöst oder protokolliert werden soll. Wenn diese Eigenschaft auf <b>true</b> festgelegt wurde, wird eine <code>ConsistencyException</code> ausgelöst.

Eigenschaft	Standardwert	Beschreibung
<code>fs.s3.consistent.metadata.autoCreate</code>	<b>true</b>	Wenn diese Eigenschaft auf <b>true</b> festgelegt wurde, ermöglicht sie die automatische Erstellung von Metadaten tabellen.
<code>fs.s3.consistent.metadata.tagVerificationEnabled</code>	<b>true</b>	Mit Amazon EMR 5.29.0 ist diese Eigenschaft standardmäßig aktiviert. Wenn diese Option aktiviert ist, verwendet EMRFS S3-ETags, um zu überprüfen, ob gelesene Objekte die neueste verfügbare Version sind. Diese Funktion ist hilfreich für read-after-update Anwendungsfälle, in denen Dateien auf S3 überschrieben werden, während derselbe Name beibehalten wird. Diese ETag-Verifizierungsfunktion funktioniert derzeit nicht mit S3 Select.
<code>fs.s3.consistent.metadata.tableName</code>	<b>EmrFSMetadata</b>	Diese Eigenschaft gibt den Namen der Metadatentabelle in DynamoDB an.
<code>fs.s3.consistent.metadata.readCapacity</code>	<b>500</b>	Diese Eigenschaft gibt die DynamoDB-Lesekapazität an, die bei Erstellung der Metadaten tabelle bereitgestellt wird.

Eigenschaft	Standardwert	Beschreibung
<code>fs.s3.consistent.metadata.write.capacity</code>	<b>100</b>	Diese Eigenschaft gibt die DynamoDB-Schreibkapazität an, die bei Erstellung der Metadaten tabelle bereitgestellt wird.
<code>fs.s3.consistent.fastList</code>	<b>true</b>	Wenn diese Eigenschaft auf <b>true</b> festgelegt wurde, verwendet sie mehrere Threads zum Auflisten eines Verzeichnisses (wenn notwendig). Damit diese Eigenschaft verwendet werden kann, muss Konsistenz aktiviert sein.
<code>fs.s3.consistent.fastList.prefetchMetadata</code>	<b>false</b>	Wenn diese Eigenschaft auf <b>true</b> festgelegt wurde, ermöglicht sie den Vorabruf von Metadaten für Verzeichnisse mit mehr als 20.000 Elementen.
<code>fs.s3.consistent.notification.CloudWatch</code>	<b>false</b>	Wenn diese Option auf gesetzt ist <b>true</b> , werden CloudWatch Metriken für FileSystem API-Aufrufe aktiviert, die aufgrund eventueller Konsistenzprobleme von Amazon S3 fehlschlagen.
<code>fs.s3.consistent.notification.SQS</code>	<b>false</b>	Wenn diese Eigenschaft auf <b>true</b> festgelegt wurde, werden Benachrichtigungen zu eventuell en Konsistenzproblemen per Push an eine Amazon-SQS-Warteschlange gesendet.

Eigenschaft	Standardwert	Beschreibung
<code>fs.s3.consistent.notification.SQS.queueName</code>	<b>EMRFS-Inconsistency- &lt;jobFlowId&gt;</b>	Wenn Sie diese Eigenschaft ändern, können Sie Ihren eigenen SQS-Warteschlangenamen für Nachrichten zu eventuellen Konsistenzproblemen von Amazon S3 angeben.
<code>fs.s3.consistent.notification.SQS.customMsg</code>	<b>none</b>	Über diese Eigenschaft können Sie benutzerdefinierte Informationen angeben, die in SQS-Nachrichten zu eventuellen Konsistenzproblemen von Amazon S3 enthalten sein sollen. Wenn für diese Eigenschaft kein Wert festgelegt ist, ist das entsprechende Feld in der Nachricht leer.
<code>fs.s3.consistent.dynamodb.endpoint</code>	<b>none</b>	Über diese Eigenschaft können Sie einen benutzerdefinierten DynamoDB-Endpunkt für Ihre Metadaten der konsistenten Ansicht angeben.
<code>fs.s3.useRequesterPaysHeader</code>	<b>false</b>	Wenn diese Eigenschaft auf <b>true</b> gesetzt ist, erlaubt sie Amazon-S3-Anfragen an Buckets mit aktivierter Bezahler-Option.

## Die EMRFS-CLI-Befehlsreferenz

Die EMRFS-CLI ist standardmäßig auf allen Cluster-Hauptknoten installiert, die mit Amazon EMR Version 3.2.1 oder höher erstellt wurden. Sie können die EMRFS CLI zum Verwalten der Metadaten für die konsistente Ansicht verwenden.

### Note

Der Befehl `emrfs` wird nur bei der VT100-Terminal-Emulation unterstützt. Er funktioniert jedoch unter Umständen auch mit anderen Terminal-Emulator-Modi.

## Übergeordneter Befehl "emrfs"

Der Top-Level-Befehl `emrfs` unterstützt die folgende Struktur.

```
emrfs [describe-metadata | set-metadata-capacity | delete-metadata | create-metadata |
 \
 list-metadata-stores | diff | delete | sync | import ] [options] [arguments]
```

Geben Sie [Optionen] mit oder ohne [Argumente] an, wie in der folgenden Tabelle beschrieben. Für [Optionen], die speziell für Unterbefehle (`describe-metadata`, `set-metadata-capacity` usw.) gelten, finden Sie nachstehend die jeweiligen Unterbefehle.

[Optionen] für `emrfs`

Option	Beschreibung	Erforderlich
<code>-a <i>AWS_ACCESS_KEY_ID</i></code>   <code>--access-key <i>AWS_ACCESS_KEY_ID</i></code>	Der AWS Zugriffsschlüssel, den Sie verwenden, um Objekte in Amazon S3 zu schreiben und einen Metadatenpeicher in DynamoDB zu erstellen oder darauf zuzugreifen. Standardmäßig ist <i>AWS_ACCESS_KEY_ID</i> auf den Zugriffsschlüssel festgelegt, der für die Cluster-Erstellung verwendet wird.	Nein
		Nein

Option	Beschreibung	Erforderlich
<code>-s <i>AWS_SECRET_ACCESS_KEY</i>   --secret-key <i>AWS_SECRET_ACCESS_KEY</i></code>	Der AWS geheime Schlüssel, der dem Zugriffsschlüssel zugeordnet ist, den Sie verwenden, um Objekte in Amazon S3 zu schreiben und einen Metadatenpeicher in DynamoDB zu erstellen oder darauf zuzugreifen. Standardmäßig ist <i>AWS_SECRET_ACCESS_KEY</i> auf den geheimen Schlüssel festgelegt, der mit dem Zugriffsschlüssel verknüpft ist, der für die Cluster-Erstellung verwendet wird.	
<code>-v   --verbose</code>	Erzeugt eine ausführliche Ausgabe.	Nein
<code>-h   --help</code>	Zeigt die Hilfenachricht für den <code>emrfs</code> -Befehl mit einer Verwendungsanweisung an.	Nein

## Unterbefehl "emrfs describe-metadata"

[Optionen] für `emrfs describe-metadata`

Option	Beschreibung	Erforderlich
<code>-m <i>METADATA_NAME</i>   --metadata-name <i>METADATA_NAME</i></code>	<i>METADATA_NAME</i> ist der Name der DynamoDB-Metadaten-tabelle. Wenn das Argument <i>METADATA_NAME</i> nicht angegeben wird, ist der Standardwert <code>EmrFSMetadata</code> .	Nein

## Example Beispiel für "emrfs describe-metadata"

Das folgende Beispiel beschreibt die Standardmetadaten-tabelle.

```
$ emrfs describe-metadata
EmrFSMetadata
  read-capacity: 400
```



```

write-capacity: 100
status: ACTIVE
approximate-item-count (6 hour delay): 12

```

## Unterbefehl `emrfs set-metadata-capacity`

[Optionen] für `emrfs set-metadata-capacity`

Option	Beschreibung	Erforderlich
-m <i>METADATA_NAME</i>   --metadata-name <i>METADATA_NAME</i>	<i>METADATA_NAME</i> ist der Name der DynamoDB-Metadaten-tabelle. Wenn das Argument <i>METADATA_NAME</i> nicht angegeben wird, ist der Standardwert <code>EmrFSMetadata</code> .	Nein
-r <i>READ_CAPACITY</i>   --read-capacity <i>READ_CAPACITY</i>	Die angeforderte Lesedurchsatzkapazität für die Metadaten-tabelle. Wenn das Argument <i>READ_CAPACITY</i> nicht angegeben wird, ist der Standardwert <code>400</code> .	Nein
-w <i>WRITE_CAPACITY</i>   --write-capacity <i>WRITE_CAPACITY</i>	Die angeforderte Schreibdurchsatzkapazität für die Metadaten-tabelle. Wenn das Argument <i>WRITE_CAPACITY</i> nicht angegeben wird, ist der Standardwert <code>100</code> .	Nein

## Example Beispiel für `emrfs set-metadata-capacity`

Im folgenden Beispiel ist für die Metadaten-tabelle `600` die Lesedurchsatzkapazität auf `150` und die Schreibkapazität auf `EmrMetadataAlt` festgelegt.

```

$ emrfs set-metadata-capacity --metadata-name EmrMetadataAlt --read-capacity 600 --
write-capacity 150
  read-capacity: 400
  write-capacity: 100
  status: UPDATING
  approximate-item-count (6 hour delay): 0

```

## Unterbefehl "emrfs delete-metadata"

[Optionen] für emrfs delete-metadata

Option	Beschreibung	Erforderlich
-m <i>METADATA_NAME</i>   --metadata-name <i>METADATA_NAME</i>	<i>METADATA_NAME</i> ist der Name der DynamoDB-Metadaten-tabelle. Wenn das Argument <i>METADATA_NAME</i> nicht angegeben wird, ist der Standardwert <code>EmrFSMetadata</code> .	Nein

Example Beispiel für "emrfs delete-metadata"

Das folgende Beispiel löscht die Standardmetadaten-tabelle.

```
$ emrfs delete-metadata
```

## Unterbefehl "emrfs create-metadata"

[Optionen] für emrfs create-metadata

Option	Beschreibung	Erforderlich
-m <i>METADATA_NAME</i>   --metadata-name <i>METADATA_NAME</i>	<i>METADATA_NAME</i> ist der Name der DynamoDB-Metadaten-tabelle. Wenn das Argument <i>METADATA_NAME</i> nicht angegeben wird, ist der Standardwert <code>EmrFSMetadata</code> .	Nein
-r <i>READ_CAPACITY</i>   --read-capacity <i>READ_CAPACITY</i>	Die angeforderte Lesedurchsatzkapazität für die Metadaten-tabelle. Wenn das Argument <i>READ_CAPACITY</i> nicht angegeben wird, ist der Standardwert <code>400</code> .	Nein
		Nein

Option	Beschreibung	Erforderlich
-w <i>WRITE_CAPACITY</i>   --write-capacity <i>WRITE_CAPACITY</i>	Die angeforderte Schreibdurchsatzkapazität für die Metadatentabelle. Wenn das Argument <i>WRITE_CAPACITY</i> nicht angegeben wird, ist der Standardwert 100.	

### Example Beispiel für "emrfs create-metadata"

Im folgenden Beispiel wird eine Metadatentabelle mit dem Namen EmrFSMetadataAlt erstellt.

```
$ emrfs create-metadata -m EmrFSMetadataAlt
Creating metadata: EmrFSMetadataAlt
EmrFSMetadataAlt
  read-capacity: 400
  write-capacity: 100
  status: ACTIVE
  approximate-item-count (6 hour delay): 0
```

### Unterbefehl emrfs list-metadata-stores

Der Unterbefehl emrfs list-metadata-stores hat keine [Optionen].

### Example L ist-metadata-stores Beispiel

Im folgenden Beispiel werden Ihre Metadatentabellen aufgelistet.

```
$ emrfs list-metadata-stores
EmrFSMetadata
```

### Unterbefehl "emrfs diff"

[Optionen] für emrfs diff

Option	Beschreibung	Erforderlich
	<i>METADATA_NAME</i> ist der Name der DynamoDB-Metadatentabelle. Wenn das Argument <i>METADATA_</i>	Nein

Option	Beschreibung	Erforderlich
<code>-m <i>METADATA_NAME</i></code>   <code>--metadata-name <i>METADATA_NAME</i></code>	<i>NAME</i> nicht angegeben wird, ist der Standardwert <code>EmrFSMetadata</code> .	
<code>s3://s3Path</code>	Der Pfad zu dem Amazon-S3-Bucket für den Vergleich mit der Metadaten-Tabelle. Die Buckets werden rekursiv synchronisiert.	Ja

### Example Beispiel für "emrfs diff"

Im folgenden Beispiel wird die Standardmetadaten-Tabelle mit einem Amazon-S3-Bucket verglichen.

```
$ emrfs diff s3://elasticmapreduce/samples/cloudfront
BOTH | MANIFEST ONLY | S3 ONLY
DIR elasticmapreduce/samples/cloudfront
DIR elasticmapreduce/samples/cloudfront/code/
DIR elasticmapreduce/samples/cloudfront/input/
DIR elasticmapreduce/samples/cloudfront/logprocessor.jar
DIR elasticmapreduce/samples/cloudfront/input/XABCD12345678.2009-05-05-14.WxYz1234
DIR elasticmapreduce/samples/cloudfront/input/XABCD12345678.2009-05-05-15.WxYz1234
DIR elasticmapreduce/samples/cloudfront/input/XABCD12345678.2009-05-05-16.WxYz1234
DIR elasticmapreduce/samples/cloudfront/input/XABCD12345678.2009-05-05-17.WxYz1234
DIR elasticmapreduce/samples/cloudfront/input/XABCD12345678.2009-05-05-18.WxYz1234
DIR elasticmapreduce/samples/cloudfront/input/XABCD12345678.2009-05-05-19.WxYz1234
DIR elasticmapreduce/samples/cloudfront/input/XABCD12345678.2009-05-05-20.WxYz1234
DIR elasticmapreduce/samples/cloudfront/code/cloudfront-loganalyzer.tgz
```

### Unterbefehl "emrfs delete"

[Optionen] für emrfs delete

Option	Beschreibung	Erforderlich
		Nein

Option	Beschreibung	Erforderlich
-m <i>METADATA_NAME</i>   --metadata-name <i>METADATA_NAME</i>	<i>METADATA_NAME</i> ist der Name der DynamoDB-Metadatentabelle. Wenn das Argument <i>METADATA_NAME</i> nicht angegeben wird, ist der Standardwert <code>EmrFSMetadata</code> .	
<i>s3://s3Path</i>	Der Pfad zum Amazon-S3-Bucket, den Sie für die konsistente Ansicht verfolgen. Die Buckets werden rekursiv synchronisiert.	Ja
-t <i>TIME</i>   --time <i>TIME</i>	Die Ablaufzeit (interpretiert mithilfe des Zeiteinheit-Arguments). Alle Metadateneinträge, die älter als das Argument <i>TIME</i> sind, werden für den angegebenen Bucket gelöscht.	
-u <i>UNIT</i>   --time-unit <i>UNIT</i>	Die Maßeinheit für die Interpretation des Zeitarguments (Nanosekunden, Mikrosekunden, Millisekunden, Sekunden, Minuten, Stunden oder Tage). Wenn kein Argument angegeben ist, lautet der Standardwert <code>days</code> .	
--read-consumption <i>READ_CONSUMPTION</i>	Die angeforderte Menge an verfügbarem Lesedurchsatz für die Operation <code>delete</code> . Wenn das Argument <i>READ_CONSUMPTION</i> nicht angegeben wird, ist der Standardwert <code>400</code> .	Nein
--write-consumption <i>WRITE_CONSUMPTION</i>	Die angeforderte Menge an verfügbarem Schreibdurchsatz für die Operation <code>delete</code> . Wenn das Argument <i>WRITE_CONSUMPTION</i> nicht angegeben wurde, ist der Standardwert <code>100</code> .	Nein

## Example Beispiel für "emrfs delete"

Im folgenden Beispiel werden alle Objekte in einem Amazon-S3-Bucket aus den verfolgenden Metadaten für die konsistente Ansicht entfernt.

```
$ emrfs delete s3://elasticmapreduce/samples/cloudfront
entries deleted: 11
```

## Unterbefehl "emrfs import"

[Optionen] für emrfs import

Option	Beschreibung	Erforderlich
-m <i>METADATA_NAME</i>   --metadata-name <i>METADATA_NAME</i>	<i>METADATA_NAME</i> ist der Name der DynamoDB-Metadaten-Tabelle. Wenn das Argument <i>METADATA_NAME</i> nicht angegeben wird, ist der Standardwert <code>EmrFSMetadata</code> .	Nein
<i>s3://s3Path</i>	Der Pfad zum Amazon-S3-Bucket, den Sie für die konsistente Ansicht verfolgen. Die Buckets werden rekursiv synchronisiert.	Ja
--read-consumption <i>READ_CONSUMPTION</i>	Die angeforderte Menge an verfügbarem Lesedurchsatz für die Operation delete. Wenn das Argument <i>READ_CONSUMPTION</i> nicht angegeben wird, ist der Standardwert 400.	Nein
--write-consumption <i>WRITE_CONSUMPTION</i>	Die angeforderte Menge an verfügbarem Schreibdurchsatz für die Operation delete. Wenn das Argument <i>WRITE_CONSUMPTION</i> nicht angegeben wurde, ist der Standardwert 100.	Nein

## Example Beispiel für "emrfs import"

Im folgenden Beispiel werden alle Objekte in einem Amazon-S3-Bucket mit den verfolgenden Metadaten für die konsistente Ansicht importiert. Alle unbekannt Schlüssel werden ignoriert.

```
$ emrfs import s3://elasticmapreduce/samples/cloudfront
```

## Unterbefehl "emrfs sync"

[Optionen] für emrfs sync

Option	Beschreibung	Erforderlich
-m <i>METADATA_NAME</i>   --metadata-name <i>METADATA_NAME</i>	<i>METADATA_NAME</i> ist der Name der DynamoDB-Metadaten-Tabelle. Wenn das Argument <i>METADATA_NAME</i> nicht angegeben wird, ist der Standardwert <code>EmrFSMetadata</code> .	Nein
<i>s3://s3Path</i>	Der Pfad zum Amazon-S3-Bucket, den Sie für die konsistente Ansicht verfolgen. Die Buckets werden rekursiv synchronisiert.	Ja
--read-consumption <i>READ_CONSUMPTION</i>	Die angeforderte Menge an verfügbarem Lesedurchsatz für die Operation delete. Wenn das Argument <i>READ_CONSUMPTION</i> nicht angegeben wird, ist der Standardwert 400.	Nein
--write-consumption <i>WRITE_CONSUMPTION</i>	Die angeforderte Menge an verfügbarem Schreibdurchsatz für die Operation delete. Wenn das Argument <i>WRITE_CONSUMPTION</i> nicht angegeben wurde, ist der Standardwert 100.	Nein

## Example Beispiel für "emrfs sync"

Im folgenden Beispiel werden alle Objekte in einem Amazon-S3-Bucket mit den verfolgenden Metadaten für die konsistente Ansicht importiert. Alle unbekannt Schlüssel werden gelöscht.

```
$ emrfs sync s3://elasticmapreduce/samples/cloudfront
Synching samples/cloudfront                0 added | 0 updated |
 0 removed | 0 unchanged
Synching samples/cloudfront/code/          1 added | 0 updated |
 0 removed | 0 unchanged
Synching samples/cloudfront/                2 added | 0 updated |
 0 removed | 0 unchanged
Synching samples/cloudfront/input/         9 added | 0 updated |
 0 removed | 0 unchanged
Done synching s3://elasticmapreduce/samples/cloudfront 9 added | 0 updated |
 1 removed | 0 unchanged
creating 3 folder key(s)
folders written: 3
```

## Unterbefehl "emrfs read-sqs"

[Optionen] für emrfs read-sqs

Option	Beschreibung	Erforderlich
<code>-q <i>QUEUE_NAME</i>   --queue-name <i>QUEUE_NAME</i></code>	<i>QUEUE_NAME</i> ist der Name der Amazon-SQS-Warteschlange, die in <code>emrfs-site.xml</code> konfiguriert ist. Der Standardwert ist <b>EMRFS-Inc consistency-&lt;jobFlowId&gt;</b> .	Ja
<code>-o <i>OUTPUT_FILE</i>   --output-file <i>OUTPUT_FILE</i></code>	<i>OUTPUT_FILE</i> ist der Pfad zur Ausgabedatei im lokalen Dateisystem des Master-Knotens. Nachrichten, die aus der Warteschlange gelesen werden, werden in diese Datei geschrieben.	Ja



## Unterbefehl "emrfs delete-sqs"

[Optionen] für emrfs delete-sqs

Option	Beschreibung	Erforderlich
<code>-q <i>QUEUE_NAME</i>   --queue-name <i>QUEUE_NAME</i></code>	<i>QUEUE_NAME</i> ist der Name der Amazon-SQS-Warteschlange, die in <code>emrfs-site.xml</code> konfiguriert ist. Der Standardwert ist <b>EMRFS-Inconsistency-&lt;jobFlowId&gt;</b> .	Ja

## Übermitteln von EMRFS-CLI-Befehlen als Schritte

Das folgende Beispiel zeigt, wie Sie das `emrfs` Hilfsprogramm auf dem Master-Knoten verwenden, indem Sie die AWS CLI OR-API nutzen und `command-runner.jar` den `emrfs` Befehl als Schritt ausführen. Das Beispiel verwendet den AWS SDK for Python (Boto3), um einem Cluster einen Schritt hinzuzufügen, der Objekte in einem Amazon S3 S3-Bucket zur standardmäßigen EMRFS-Metadatentabelle hinzufügt.

```
import boto3
from botocore.exceptions import ClientError

def add_emrfs_step(command, bucket_url, cluster_id, emr_client):
    """
    Add an EMRFS command as a job flow step to an existing cluster.

    :param command: The EMRFS command to run.
    :param bucket_url: The URL of a bucket that contains tracking metadata.
    :param cluster_id: The ID of the cluster to update.
    :param emr_client: The Boto3 Amazon EMR client object.
    :return: The ID of the added job flow step. Status can be tracked by calling
             the emr_client.describe_step() function.
    """
    job_flow_step = {
        "Name": "Example EMRFS Command Step",
        "ActionOnFailure": "CONTINUE",
        "HadoopJarStep": {
            "Jar": "command-runner.jar",
```

```
        "Args": ["/usr/bin/emrfs", command, bucket_url],
    },
}

try:
    response = emr_client.add_job_flow_steps(
        JobFlowId=cluster_id, Steps=[job_flow_step]
    )
    step_id = response["StepIds"][0]
    print(f"Added step {step_id} to cluster {cluster_id}.")
except ClientError:
    print(f"Couldn't add a step to cluster {cluster_id}.")
    raise
else:
    return step_id

def usage_demo():
    emr_client = boto3.client("emr")
    # Assumes the first waiting cluster has EMRFS enabled and has created metadata
    # with the default name of 'EmrFSMetadata'.
    cluster = emr_client.list_clusters(ClusterStates=["WAITING"])["Clusters"][0]
    add_emrfs_step(
        "sync", "s3://elasticmapreduce/samples/cloudfront", cluster["Id"], emr_client
    )

if __name__ == "__main__":
    usage_demo()
```

Sie können den zurückgegebenen Wert `step_id` verwenden, um die Protokolle auf das Ergebnis der Operation zu prüfen.

## Zugriff auf EMRFS-Daten in Amazon S3 genehmigen

Standardmäßig bestimmt die EMR-Rolle für EC2 die Berechtigungen für den Zugriff auf EMRFS-Daten in Amazon S3. Die IAM-Richtlinien, die dieser Rolle zugeordnet werden, gelten unabhängig von dem oder der Gruppe, die die Anforderung über EMRFS erstellen. Der Standardwert ist `EMR_EC2_DefaultRole`. Weitere Informationen finden Sie unter [Servicerolle für EC2-Cluster-Instances \(EC2-Instance-Profil\)](#).

Ab Amazon-EMR-Version 5.10.0 können Sie mittels einer Sicherheitskonfiguration IAM-Rollen für EMRFS angeben. Auf diese Weise können Sie Berechtigungen für EMRFS-Anforderungen an Amazon S3 für Cluster mit mehreren Benutzern anpassen. Sie können verschiedene IAM-Rollen für verschiedene Benutzer und Gruppen und für verschiedene Amazon-S3-Bucket-Speicherorte angeben, basierend auf dem Präfix in Amazon S3. Wenn EMRFS eine Anforderung an Amazon S3 stellt, die mit Benutzern, Gruppen oder Standorten, die Sie angeben, übereinstimmt, verwendet der Cluster die entsprechende Rolle und nicht die EMR-Rolle für EC2. Weitere Informationen finden Sie unter [Konfigurieren von IAM-Rollen für EMRFS-Anforderungen an Amazon S3](#).

Wenn die Anforderungen Ihrer Amazon-EMR-Lösung über die Möglichkeiten hinausgehen, die IAM-Rollen für EMRFS bereitstellen, können Sie alternativ eine benutzerdefinierte Anmeldeinformationsanbieter-Klasse definieren, mit der Sie den Zugriff auf EMRFS-Daten in Amazon S3 anpassen können.

## Erstellen eines benutzerdefinierten Anmeldeinformationsanbieters für EMRFS-Daten in Amazon S3

[Um einen Anbieter für benutzerdefinierte Anmeldeinformationen zu erstellen, implementieren Sie die Klassen AWSCredentialsProvider und Hadoop Configurable.](#)

Eine ausführliche Erläuterung dieses Ansatzes finden Sie im AWS Big Data-Blog unter [Daten aus einem anderen AWS Konto mit EMRFS sicher analysieren](#). Der Blogbeitrag enthält ein Tutorial, das Sie durch den Prozess führtend-to-end, von der Erstellung von IAM-Rollen bis hin zum Starten des Clusters. Er enthält auch ein Java-Codebeispiel zum Implementieren der benutzerdefinierten Anmeldeinformationsanbieter-Klasse.

Die grundlegenden Schritte sind wie folgt:

So definieren Sie einen benutzerdefinierten Anmeldeinformationsanbieter

1. Erstellen Sie eine benutzerdefinierte Anmeldeinformationsanbieter-Klasse als JAR-Datei.
2. Führen Sie ein Skript als Bootstrap-Aktion zum Kopieren der JAR-Datei mit dem benutzerdefinierten Anmeldeinformationsanbieter in `/usr/share/aws/emr/emrfs/auxlib` im Master-Knoten des Clusters aus. Weitere Informationen zu Bootstrap-Aktionen finden Sie unter [\(Optional\) Erstellen von Bootstrap-Aktionen zum Installieren zusätzlicher Software](#).
3. Passen Sie die `emrfs-site`-Klassifizierung an, um die in der JAR-Datei implementierte Klasse anzugeben. Weitere Informationen zur Angabe von Konfigurationsobjekten für die

Anpassung von Anwendungen finden Sie unter [Konfigurieren von Anwendungen](#) im Amazon-EMR-Versionshandbuch.

Das folgende Beispiel zeigt einen `create-cluster`-Befehl, der einen Hive-Cluster mit üblichen Konfigurationsparametern startet und außerdem Folgendes umfasst:

- Eine Bootstrap-Aktion, die das Skript `copy_jar_file.sh` ausführt, das in `mybucket` in Amazon S3 gespeichert wird.
- Eine `emrfs-site`-Klassifizierung, in der ein in der JAR-Datei als benutzerdefinierter Anmeldeinformationsanbieter als `MyCustomCredentialsProvider` definiert wird.

### Note

Linux-Zeilenfortsetzungszeichen (`\`) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (`^`).

```
aws emr create-cluster --applications Name=Hive \
--bootstrap-actions '[{"Path":"s3://mybucket/copy_jar_file.sh", "Name":"Custom
action"}]' \
--ec2-attributes '{"KeyName":"MyKeyPair", "InstanceProfile":"EMR_EC2_DefaultRole", \
"SubnetId":"subnet-xxxxxxx", "EmrManagedSlaveSecurityGroup":"sg-xxxxxxx", \
"EmrManagedMasterSecurityGroup":"sg-xxxxxxx"}' \
--service-role EMR_DefaultRole_V2 --enable-debugging --release-label emr-7.1.0 \
--log-uri 's3n://my-emr-log-bucket/' --name 'test-awscredentialsprovider-emrfs' \
--instance-type=m5.xlarge --instance-count 3 \
--configurations '[{"Classification":"emrfs-site", \
"Properties":
{"fs.s3.customAWSCredentialsProvider":"MyAWSCredentialsProviderWithUri"}, \
"Configurations":[]}]'
```

## Verwaltung des Standardendpunkts AWS Security Token Service

EMRFS verwendet das AWS Security Token Service (STS), um temporäre Sicherheitsanmeldeinformationen für den Zugriff auf Ihre Ressourcen abzurufen. AWS Frühere Amazon EMR-Release-Versionen senden alle AWS STS Anfragen an einen einzigen globalen

Endpunkt unter `https://sts.amazonaws.com`. Amazon EMR veröffentlicht die Versionen 5.31.0 und 6.1.0 und höher und sendet stattdessen Anfragen an regionale Endpunkte. AWS STS Dies reduziert die Latenz und verbessert die Gültigkeit von Sitzungstoken. Weitere Informationen zu AWS STS Endpunkten finden Sie im Benutzerhandbuch unter [Verwaltung AWS STS in einer AWS Region](#). AWS Identity and Access Management

Wenn Sie die Amazon-EMR-Release-Versionen 5.31.0 und 6.1.0 und höher verwenden, können Sie den AWS STS -Standard-Endpunkt überschreiben. Dazu müssen Sie die `fs.s3.sts.endpoint`-Eigenschaft in Ihrer `emrfs-site`-Konfiguration ändern.

Im folgenden AWS CLI Beispiel wird der von EMRFS verwendete AWS STS Standardendpunkt auf den globalen Endpunkt festgelegt.

```
aws emr create-cluster --release-label <emr-5.33.0> --instance-type m5.xlarge \  
--emrfs Args=[fs.s3.sts.endpoint=https://sts.amazonaws.com]
```

#### Note

Linux-Zeilenfortsetzungszeichen (`\`) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (`^`).

Alternativ können Sie mithilfe des folgenden Beispiels eine JSON-Konfigurationsdatei erstellen und sie mit dem `--configurations`-Argument von `emr create-cluster` angeben. Weitere Informationen zur Verwendung `--configurations`, finden Sie unter [AWS CLI in der Command-Befehlsreferenz](#).

```
[  
  {  
    "classification": "emrfs-site",  
    "properties": {  
      "fs.s3.sts.endpoint": "https://sts.amazonaws.com"  
    }  
  }  
]
```

# Angeben der Amazon-S3-Verschlüsselung mit EMRFS-Eigenschaften

## Important

Ab Amazon-EMR-Version 4.8.0 können Sie Sicherheitskonfigurationen verwenden, um Verschlüsselungseinstellungen einfacher und mit mehr Optionen einzurichten. Wir empfehlen die Verwendung von Sicherheitskonfigurationen. Weitere Informationen finden Sie unter [Datenverschlüsselung konfigurieren](#). Die in diesem Abschnitt beschriebenen Konsolenbefehle sind für Versionen vor 4.8.0 verfügbar. Wenn Sie die AWS CLI Amazon S3 S3-Verschlüsselung sowohl in der Cluster-Konfiguration als auch in einer Sicherheitskonfiguration in nachfolgenden Versionen verwenden, überschreibt die Sicherheitskonfiguration die Cluster-Konfiguration.


Wenn Sie einen Cluster erstellen, können Sie serverseitige Verschlüsselung (SSE) oder clientseitige Verschlüsselung (CSE) für EMRFS-Daten in Amazon S3 mithilfe der Konsole oder mithilfe von `emrfs-site` Klassifizierungseigenschaften über das oder das EMR-SDK angeben. AWS CLI SSE und CSE in Amazon S3 schließen sich gegenseitig aus. Sie können jede der beiden Optionen, aber nicht beide auswählen.

AWS CLI Anweisungen finden Sie unten im entsprechenden Abschnitt für Ihren Verschlüsselungstyp.

Um EMRFS-Verschlüsselungsoptionen anzugeben, verwenden Sie AWS Management Console

1. Navigieren Sie zur neuen Amazon-EMR-Konsole und wählen Sie in der Seitennavigation die Option [Zur alten Konsole wechseln](#) aus. Weitere Informationen darüber, was Sie erwartet, wenn Sie zur alten Konsole wechseln, finden Sie unter [Verwenden der alten Konsole](#).
2. Wählen Sie **Create Cluster** (Cluster erstellen) und **Go to advanced options** (Zu erweiterten Optionen) aus.
3. Wählen Sie **Release (Version) 4.7.2** oder früher aus.
4. Wählen Sie weitere Optionen für **Software and Steps** (Software und Steps) aus wie für Ihre Anwendung erforderlich. Wählen Sie anschließend **Next** (Weiter) aus.
5. Wählen Sie in den Bereichen **Hardware** (Hardware) und **General Cluster Settings** (Allgemeine Cluster-Einstellungen) Einstellungen aus wie für Ihre Anwendung erforderlich.

- Wählen Sie im Bereich Security (Sicherheit) in Authentication and encryption (Authentifizierung und Verschlüsselung) die zu verwendende Option S3 Encryption (with EMRFS) (S3-Verschlüsselung (mit EMRFS)) aus.

 Note

Serverseitige S3-Verschlüsselung mit KMS-Schlüsselverwaltung (SSE-KMS) steht für Amazon-EMR-Version 4.4 oder früher nicht zur Verfügung.

- Wenn Sie eine Option wählen, die AWS -Schlüsselverwaltung verwendet, müssen Sie eine AWS -KMS-Schlüssel-ID auswählen. Weitere Informationen finden Sie unter [Wird AWS KMS keys für die EMRFS-Verschlüsselung verwendet](#).
  - Wenn Sie S3 client-side encryption with custom materials provider (Clientseitige S3-Verschlüsselung mit benutzerdefiniertem Materialanbieter) auswählen, müssen Sie einen Wert in Class name (Klassenname) und JAR location (JAR-Speicherort) angeben. Weitere Informationen finden Sie unter [Clientseitige Verschlüsselung für Amazon S3](#).
- Wählen Sie weitere Optionen wie für Ihre Anwendung erforderlich aus. Wählen Sie anschließend Create Cluster (Cluster erstellen) aus.

## Wird AWS KMS keys für die EMRFS-Verschlüsselung verwendet

Der AWS KMS Verschlüsselungsschlüssel muss in derselben Region wie Ihre Amazon EMR-Cluster-Instance und die mit EMRFS verwendeten Amazon S3 S3-Buckets erstellt werden. Wenn sich der von Ihnen angegebene Schlüssel in einem anderen Konto befindet als dem, das Sie zur Konfiguration eines Clusters verwenden, müssen Sie den Schlüssel mit seinem ARN angeben.

Die Rolle für das Amazon-EC2-Instance-Profil muss über die Berechtigung zur Nutzung des von Ihnen angegebenen KMS-Schlüssels verfügen. Die Standardrolle für das Instance-Profil in Amazon EMR ist `EMR_EC2_DefaultRole`. Wenn Sie eine andere Rolle für das Instance-Profil oder IAM-Rollen für EMRFS-Anfragen an Amazon S3 verwenden, stellen Sie sicher, dass jede Rolle je nach Bedarf als Schlüsselbenutzer hinzugefügt wird. So erhält die Rolle die Berechtigung, den KMS-Schlüssel zu verwenden. Weitere Informationen finden Sie unter [Nutzung von Schlüsselrichtlinien](#) im AWS Key Management Service -Entwicklerhandbuch und [Konfigurieren von IAM-Rollen für EMRFS-Anfragen an Amazon S3](#).

Sie können das verwenden AWS Management Console , um Ihr Instance-Profil oder EC2-Instance-Profil zur Liste der Schlüsselbenutzer für den angegebenen KMS-Schlüssel hinzuzufügen, oder Sie können das oder ein AWS SDK verwenden, um eine AWS CLI entsprechende Schlüsselrichtlinie anzuhängen.

Hinweis: Amazon EMR unterstützt nur [symmetrische KMS-Schlüssel](#). Sie können keinen [asymmetrischen KMS-Schlüssel](#) verwenden, um Data-at-Rest in einem Amazon-EMR-Cluster zu verschlüsseln. Wie Sie feststellen, ob ein KMS-Schlüssel symmetrisch oder asymmetrisch ist, erfahren Sie unter [Erkennen symmetrischer und asymmetrischer KMS-Schlüssel](#).

Im folgenden Verfahren wird beschrieben, wie Sie das Amazon-EMR-Instance-Profil mithilfe der `EMR_EC2_DefaultRole` als Schlüsselbenutzer mit AWS Management Console hinzufügen. Dabei wird davon ausgegangen, dass Sie bereits einen KMS-Schlüssel erstellt haben. Weitere Informationen über die Erstellung eines neuen KMS-Schlüssels finden Sie unter [Erstellen von Schlüsseln](#) im AWS Key Management Service -Entwicklerhandbuch.

So fügen Sie das EC2-Instance-Profil für Amazon EMR zur Liste der Verschlüsselungsschlüssel-Benutzer hinzu

1. Melden Sie sich bei der AWS Key Management Service (AWS KMS) -Konsole an AWS Management Console und öffnen Sie sie unter <https://console.aws.amazon.com/kms>.
2. Um das zu ändern AWS-Region, verwenden Sie die Regionsauswahl in der oberen rechten Ecke der Seite.
3. Wählen Sie den Alias des zu ändernden KMS-Schlüssels aus.
4. Wählen Sie auf der Seite mit den Schlüsseldetails unter Key Users (Schlüsselbenutzer( die Option Add (Hinzufügen) aus.
5. Wählen Sie die entsprechende Rolle im Dialogfeld Add key users (Schlüsselbenutzer hinzufügen) aus. Der Name der Standardrolle lautet `EMR_EC2_DefaultRole`.
6. Wählen Sie Hinzufügen aus.

## Serverseitige Verschlüsselung im Amazon S3

Wenn Sie die Amazon-S3-Verschlüsselung einrichten, verschlüsselt Amazon S3 die Daten auf der Objektebene, während die Daten auf den Datenträger geschrieben werden, und entschlüsselt sie, wenn auf sie zugegriffen wird. Weitere Informationen über SSE finden Sie unter [Schutz von Daten durch serverseitige Verschlüsselung](#) im Amazon Simple Storage Service User Guide.



Wenn Sie SSE in Amazon EMR einrichten, haben Sie die Wahl zwischen zwei verschiedenen Systemen für die Schlüsselverwaltung:

- SSE-S3 – Hierbei verwaltet Amazon S3 die Aktivierungsschlüssel für Sie.
- SSE-KMS — Sie verwenden eine AWS KMS key , um Richtlinien einzurichten, die für Amazon EMR geeignet sind. Weitere Informationen zu den wichtigsten Anforderungen für Amazon EMR finden Sie unter [AWS KMS keys Zur Verschlüsselung verwenden](#).

SSE mit vom Kunden bereitgestellten Schlüsseln (SSE-C) ist für Amazon EMR nicht verfügbar.

Um einen Cluster mit aktiviertem SSE-S3 zu erstellen, verwenden Sie AWS CLI

- Geben Sie den folgenden Befehl ein:

```
aws emr create-cluster --release-label emr-4.7.2 or earlier \  
--instance-count 3 --instance-type m5.xlarge --emrfs Encryption=ServerSide
```

Sie können SSE-S3 auch aktivieren, indem Sie `fs.s3.enableServerSideEncryption` in den Eigenschaften auf `true` gesetzt. Lesen Sie dazu das Beispiel für SSE-KMS unten, und lassen Sie die Eigenschaft für Key-ID weg.

Um einen Cluster mit aktiviertem SSE-KMS zu erstellen, verwenden Sie AWS CLI

#### Note

SSE-KMS steht nur in Amazon-EMR-Version 4.5.0 und höher zur Verfügung.

- *Geben Sie den folgenden AWS CLI Befehl ein, um einen Cluster mit SSE-KMS zu erstellen, wobei KeyID beispielsweise AWS KMS key `a4567b8-9900-12ab-1234-123a45678901` ist:*

```
aws emr create-cluster --release-label emr-4.7.2 or earlier --instance-count 3 \  
--instance-type m5.xlarge --use-default-roles \  
--emrfs Encryption=ServerSide,Args=[fs.s3.serverSideEncryption.kms.keyId=keyId]
```

--ODER--

Geben Sie den folgenden AWS CLI Befehl unter Verwendung der Klassifizierung ein und stellen Sie eine JSON-Konfigurationsdatei mit Inhalten bereit, wie im Beispiel unten dargestellt: `emrfs-site myConfig.json`

```
aws emr create-cluster --release-label emr-4.7.2 or earlier --instance-count 3
--instance-type m5.xlarge --applications Name=Hadoop --configurations file://
myConfig.json --use-default-roles
```

Beispielinhalt von `myConfig.json`:

```
[
  {
    "Classification": "emrfs-site",
    "Properties": {
      "fs.s3.enableServerSideEncryption": "true",
      "fs.s3.serverSideEncryption.kms.keyId": "a4567b8-9900-12ab-1234-123a45678901"
    }
  }
]
```

## Konfigurationseigenschaften für SSE-S3 und SSE-KMS

Diese Eigenschaften können unter Verwendung der `emrfs-site`- Konfigurationsklassifikation konfiguriert werden. SSE-KMS steht nur in Amazon-EMR-Version 4.5.0 und höher zur Verfügung.

Eigenschaft	Standardwert	Beschreibung
<code>fs.s3.enableServerSideEncryption</code>	<b>false</b>	Wenn dies auf <b>true</b> festgelegt wurde, werden in Amazon S3 gespeicherte Objekte mittels serverseitiger Verschlüsselung verschlüsselt. Wenn kein Schlüssel ausgewählt wurde, wird SSE-S3 verwendet.

Eigenschaft	Standardwert	Beschreibung
<code>fs.s3.serverSideEncryption.kms.keyId</code>	n/a	Gibt eine AWS KMS Schlüssel-ID oder einen ARN an. Wenn ein Schlüssel ausgewählt wurde, wird SSE-KMS verwendet.

## Clientseitige Verschlüsselung für Amazon S3

Mit Amazon S3 bei der clientseitigen Verschlüsselung erfolgt der Amazon-S3-Ver- und Entschlüsselungsvorgang im EMRFS-Client auf Ihrem EMR-Cluster. Objekte werden vor dem Hochladen nach Amazon S3 verschlüsselt und nach dem Herunterladen entschlüsselt. Der von Ihnen festgelegte Anbieter stellt den vom Client verwendeten Verschlüsselungsschlüssel bereit. Der Client kann vom AWS KMS bereitgestellte Schlüssel (CSE-KMS) oder eine benutzerdefinierte Java-Klasse verwenden, die den clientseitigen Root-Schlüssel (CSE-C) bereitstellt. Die Verschlüsselungseigenschaften unterscheiden sich geringfügig zwischen CSE-KMS und CSE-C, abhängig vom festgelegten Anbieter und von den Metadaten des Objekts, das entschlüsselt oder verschlüsselt werden soll. Weitere Informationen zu diesen Unterschieden finden Sie unter [Schützen von Daten durch clientseitige Verschlüsselung](#) im Entwicklerhandbuch von Amazon Simple Storage Service.

### Note

Amazon S3 CSE stellt nur sicher, dass EMRFS-Daten, die mit Amazon S3 ausgetauscht werden, verschlüsselt sind. Nicht alle Daten auf den Cluster-Instance-Volumes werden verschlüsselt. Da Hue EMRFS nicht verwendet, werden darüber hinaus Objekte, die vom Hue-S3-Dateibrowser in Amazon S3 geschrieben werden, nicht verschlüsselt.

Um CSE-KMS für EMRFS-Daten in Amazon S3 zu spezifizieren, verwenden Sie den AWS CLI

- Geben Sie den folgenden Befehl ein und ersetzen Sie *MyKMSKeyID* durch die Schlüssel-ID oder den ARN des zu verwendenden KMS CMK:

```
aws emr create-cluster --release-label emr-4.7.2 or earlier
```

```
--emrfs Encryption=ClientSide,ProviderType=KMS,KMSKeyId=MyKMSKeyId
```

## Erstellen eines benutzerdefinierten Schlüsselanbieters

Abhängig von der Art der Verschlüsselung, die Sie bei der Erstellung eines benutzerdefinierten Schlüsselanbieters verwenden, muss die Anwendung auch unterschiedliche Schnittstellen implementieren. EncryptionMaterialsProvider Beide Schnittstellen sind im AWS SDK for Java Version 1.11.0 und höher verfügbar.

- Um die Amazon S3 S3-Verschlüsselung zu implementieren, verwenden Sie das Modell [com.amazonaws.services.s3.model. EncryptionMaterialsProvider Schnittstelle](#).
- Verwenden Sie die Datei [com.amazonaws.services.elasticmapreduce.spi.security, um die lokale Festplattenverschlüsselung zu implementieren. EncryptionMaterialsProvider Schnittstelle](#).

Sie können jede Strategie verwenden, um Verschlüsselungsmaterial für die Implementierung bereitzustellen. Sie können sich beispielsweise dafür entscheiden, statisches Verschlüsselungsmaterial bereitzustellen oder es in ein komplexeres Schlüsselverwaltungssystem zu integrieren.

Wenn Sie die Amazon S3 S3-Verschlüsselung verwenden, müssen Sie die Verschlüsselungsalgorithmen AES/GCM/ NoPadding für benutzerdefinierte Verschlüsselungsmaterialien verwenden.

Wenn Sie die lokale Festplattenverschlüsselung verwenden, variiert der Verschlüsselungsalgorithmus, der für benutzerdefinierte Verschlüsselungsmaterialien verwendet werden soll, je nach EMR-Version. Für Amazon EMR 7.0.0 und niedriger müssen Sie AES/GCM/ verwenden. NoPadding Für Amazon EMR 7.1.0 und höher müssen Sie AES verwenden.

Die EncryptionMaterialsProvider Klasse ruft Verschlüsselungsmaterial anhand des Verschlüsselungskontextes ab. Amazon EMR aktualisiert die Verschlüsselungskontextinformationen während der Laufzeit. So wird der Aufrufer bei der Entscheidung unterstützt, welche korrekten Verschlüsselungsmaterialien zurückzugeben sind.

Example Beispiel: Verwenden eines benutzerdefinierten Schlüsselanbieters für die Amazon-S3-Verschlüsselung mit EMRFS

Wenn Amazon EMR die Verschlüsselungsmaterialien von der EncryptionMaterialsProvider Klasse abrufen, um die Verschlüsselung durchzuführen, füllt EMRFS optional das Argument

MaterialsDescription mit zwei Feldern auf: dem Amazon S3 S3-URI für das Objekt und dem des Clusters, die JobFlowId von der Klasse verwendet werden können, um Verschlüsselungsmaterialien selektiv zurückzugeben. EncryptionMaterialsProvider

Beispielsweise kann der Anbieter unterschiedliche Schlüssel für unterschiedliche Amazon-S3-URI-Präfixe zurückgeben. Es ist die Beschreibung der zurückgegebenen Verschlüsselungsmaterialien, die schließlich mit dem Amazon-S3-Objekt gespeichert wird, und nicht der materialsDescription-Wert, der von EMRFS generiert und an den Anbieter weitergeleitet wird. Beim Entschlüsseln eines Amazon S3 S3-Objekts wird die Beschreibung des Verschlüsselungsmaterials an die EncryptionMaterialsProvider Klasse übergeben, sodass sie wiederum selektiv den passenden Schlüssel zur Entschlüsselung des Objekts zurückgeben kann.

Eine EncryptionMaterialsProvider Referenzimplementierung finden Sie weiter unten. Ein weiterer kundenspezifischer Anbieter, [EMRFSRSA EncryptionMaterialsProvider](#), ist bei erhältlich. GitHub

```
import com.amazonaws.services.s3.model.EncryptionMaterials;
import com.amazonaws.services.s3.model.EncryptionMaterialsProvider;
import com.amazonaws.services.s3.model.KMSEncryptionMaterials;
import org.apache.hadoop.conf.Configurable;
import org.apache.hadoop.conf.Configuration;

import java.util.Map;

/**
 * Provides KMSEncryptionMaterials according to Configuration
 */
public class MyEncryptionMaterialsProviders implements EncryptionMaterialsProvider,
Configurable{
    private Configuration conf;
    private String kmsKeyId;
    private EncryptionMaterials encryptionMaterials;

    private void init() {
        this.kmsKeyId = conf.get("my.kms.key.id");
        this.encryptionMaterials = new KMSEncryptionMaterials(kmsKeyId);
    }

    @Override
    public void setConf(Configuration conf) {
        this.conf = conf;
        init();
    }
}
```

```

@Override
public Configuration getConf() {
    return this.conf;
}

@Override
public void refresh() {

}

@Override
public EncryptionMaterials getEncryptionMaterials(Map<String, String>
materialsDescription) {
    return this.encryptionMaterials;
}

@Override
public EncryptionMaterials getEncryptionMaterials() {
    return this.encryptionMaterials;
}
}

```

## Spezifizieren Sie einen Anbieter für benutzerdefinierte Materialien mithilfe der AWS CLI

Um die AWS CLI zu verwenden, übergeben Sie die Argumente `Encryption`, `ProviderType`, `CustomProviderClass` und `CustomProviderLocation` an die Option `emrfs`.

```

aws emr create-cluster --instance-type m5.xlarge --release-label emr-4.7.2 or earlier
--emrfs Encryption=ClientSide,ProviderType=Custom,CustomProviderLocation=s3://
mybucket/myfolder/provider.jar,CustomProviderClass=classname

```

Das Festlegen von `Encryption` auf `ClientSide` aktiviert die clientseitige Verschlüsselung. `CustomProviderClass` ist der Name des `EncryptionMaterialsProvider`-Objekts und `CustomProviderLocation` ist der lokale oder Amazon-S3-Speicherort, aus dem Amazon EMR `CustomProviderClass` zu jedem Knoten im Cluster kopiert und in den Klassenpfad platziert.

## Angabe eines benutzerdefinierten Materialanbieters unter Verwendung eines SDK

Um ein SDK zu verwenden, können Sie die Eigenschaft `fs.s3.cse.encryptionMaterialsProvider.uri` auf das Herunterladen der

benutzerdefinierten EncryptionMaterialsProvider-Klasse festlegen, die Sie in Amazon S3 auf jedem Knoten Ihres Clusters speichern. Sie konfigurieren dies in der Datei `emrfs-site.xml` zusammen mit der aktivierten CSE sowie dem richtigen Speicherort des benutzerdefinierten Anbieters.

Bei der AWS SDK for Java Verwendung `RunJobFlowRequest` könnte Ihr Code beispielsweise wie folgt aussehen:

```
<snip>
    Map<String,String> emrfsProperties = new HashMap<String,String>();
        emrfsProperties.put("fs.s3.cse.encryptionMaterialsProvider.uri","s3://mybucket/
MyCustomEncryptionMaterialsProvider.jar");
        emrfsProperties.put("fs.s3.cse.enabled","true");
        emrfsProperties.put("fs.s3.consistent","true");

    emrfsProperties.put("fs.s3.cse.encryptionMaterialsProvider","full.class.name.of.EncryptionMate

    Configuration myEmrfsConfig = new Configuration()
        .withClassification("emrfs-site")
        .withProperties(emrfsProperties);

    RunJobFlowRequest request = new RunJobFlowRequest()
        .withName("Custom EncryptionMaterialsProvider")
        .withReleaseLabel("emr-7.1.0")
        .withApplications(myApp)
        .withConfigurations(myEmrfsConfig)
        .withServiceRole("EMR_DefaultRole_V2")
        .withJobFlowRole("EMR_EC2_DefaultRole")
        .withLogUri("s3://myLogUri/")
        .withInstances(new JobFlowInstancesConfig()
            .withEc2KeyName("myEc2Key")
            .withInstanceCount(2)
            .withKeepJobFlowAliveWhenNoSteps(true)
            .withMasterInstanceType("m5.xlarge")
            .withSlaveInstanceType("m5.xlarge")
        );

    RunJobFlowResult result = emr.runJobFlow(request);
</snip>
```

## EncryptionMaterialsProvider benutzerdefiniert mit Argumenten

Möglicherweise müssen Sie Argumente direkt dem Anbieter übergeben. Zu diesem Zweck können Sie die `emrfs-site`-Konfigurationsklassifizierung mit benutzerdefinierten Argumente als Eigenschaften verwenden. Eine Beispielkonfiguration ist nachfolgend gezeigt, die als Datei gespeichert wird, `myConfig.json`:

```
[
  {
    "Classification": "emrfs-site",
    "Properties": {
      "myProvider.arg1": "value1",
      "myProvider.arg2": "value2"
    }
  }
]
```

Mit dem `create-cluster` Befehl von können Sie die `--configurations` Option verwenden AWS CLI, um die Datei wie unten gezeigt anzugeben:

```
aws emr create-cluster --release-label emr-7.1.0 --instance-type m5.xlarge
--instance-count 2 --configurations file://myConfig.json --emrfs
Encryption=ClientSide,CustomProviderLocation=s3://mybucket/myfolder/
myprovider.jar,CustomProviderClass=classname
```

## Konfiguration der EMRFS-S3EC-V2-Unterstützung

S3-Java-SDK-Versionen (1.11.837 und höher) unterstützen den Verschlüsselungsclient Version 2 (S3EC V2) mit verschiedenen Sicherheitsverbesserungen. Weitere Informationen finden Sie im S3-Blogbeitrag: [Updates für den Amazon S3 Encryption Client](#). Weitere Informationen finden Sie im AWS SDK for Java Developer Guide unter [Migration des Amazon S3 S3-Verschlüsselungsclients](#).

Der Encryption Client V1 ist aus Gründen der Abwärtskompatibilität weiterhin im SDK verfügbar. Standardmäßig verwendet EMRFS S3EC V1 zum Verschlüsseln und Entschlüsseln von S3-Objekten, wenn CSE aktiviert ist.

Mit S3EC V2 verschlüsselte S3-Objekte können nicht von EMRFS auf einem EMR-Cluster entschlüsselt werden, dessen Release-Version älter als `emr-5.31.0` ist (`emr-5.30.1` und früher, `emr-6.1.0` und früher).



## Example Konfigurieren Sie EMRFS für die Verwendung von S3EC V2

Um EMRFS für die Verwendung von S3EC V2 zu konfigurieren, fügen Sie die folgende Konfiguration hinzu:

```
{
  "Classification": "emrfs-site",
  "Properties": {
    "fs.s3.cse.encryptionV2.enabled": "true"
  }
}
```

## emrfs-site.xml-Eigenschaften für die Amazon-S3-clientseitige Verschlüsselung

Eigenschaft	Standardwert	Beschreibung
<code>fs.s3.cse.enabled</code>	<b>false</b>	Wenn dies auf <b>true</b> festgelegt ist, werden in Amazon S3 gespeicherte EMRFS-Objekte mittels clientseitiger Verschlüsselung verschlüsselt.
<code>fs.s3.cse.encryptionV2.enabled</code>	<b>false</b>	Wenn diese Option auf <code>true</code> gesetzt ist, verwendet EMRFS den S3-Verschlüsselungsclient Version 2 zum Verschlüsseln und Entschlüsseln von Objekten auf S3. Verfügbar in der EMR-Version 5.31.0 und höher verfügbar.
<code>fs.s3.cse.encryptionMaterialsProvider.uri</code>	<b>N/A</b>	Gilt bei der Verwendung benutzerdefinierter Verschlüsselungsmaterialien. Die Amazon-S3-URI, in der sich die JAR mit dem <code>EncryptionMaterial</code>

Eigenschaft	Standardwert	Beschreibung
		<p>sProvider befindet. Wenn Sie diesen URI bereitstellen, Amazon EMR automatisch die JAR-Datei auf alle Knoten im Cluster herunter.</p>
<code>fs.s3.cse.encryptionMaterialsProvider</code>	<b>N/A</b>	<p>Der <code>EncryptionMaterial</code> sProvider -Klassenpfad, der bei der clientseitigen Verschlüsselung verwendet wird. Bei der Verwendung von CSE-KMS geben <code>com.amazon.ws.emr.hadoop.fs.cse.KMSEncryptionMaterialsProvider</code> an.</p>
<code>fs.s3.cse.materialsDescription.enabled</code>	<b>false</b>	<p>Wenn auf <code>gesetztrue</code>, füllt die <code>MaterialsDescription</code> verschlüsselter Objekte mit dem Amazon S3 S3-URI für das Objekt und den <code>JobFlowId</code>. Bei der Verwendung benutzerdefinierter Verschlüsselungsmaterialien auf <code>true</code> setzen.</p>
<code>fs.s3.cse.kms.keyId</code>	<b>N/A</b>	<p>Gilt bei der Verwendung von CSE-KMS. Der Wert des <code>KeyId</code>, ARN oder Alias des KMS-Schlüssels, der für die Verschlüsselung verwendet wird.</p>

Eigenschaft	Standardwert	Beschreibung
fs.s3.cse.cryptoStorageMode	<b>ObjectMetadata</b>	<p>Der Amazon-S3-Speichermodus. Standardmäßig wird die Beschreibung der Verschlüsselungsdaten in den Objektmetadaten gespeichert. Sie können die Beschreibung auch in einer Anweisungsdatei speichern. Gültige Werte sind ObjectMetadata und InstructionFile. Weitere Informationen finden Sie unter <a href="#">Clientseitige Datenverschlüsselung mit dem AWS SDK for Java und Amazon S3</a>.</p>

# CloudWatch Amazon-Vertreter

Der CloudWatch Amazon-Agent auf Amazon EMR ist ein Tool, das die Amazon EC2-Instances in Ihrem EMR-Cluster überwachen kann. Sie können die Metriken, die Sie mit dem Agenten sammeln, speichern und anzeigen. CloudWatch CloudWatch Weitere Informationen über den CloudWatch Agenten finden Sie im [CloudWatch Amazon-Benutzerhandbuch](#).

## Note

Es fallen zusätzliche Gebühren an, wenn Sie andere verwenden, AWS-Services um CloudWatch Amazon-Agentenmetriken zu veröffentlichen, abzufragen oder einzusehen. Weitere Preisinformationen finden Sie auf den folgenden Seiten.

- [CloudWatch Amazon-Preisgestaltung](#)
- [Preise für Amazon Managed Service für Prometheus](#)
- [Amazon Managed Grafana-Preise](#)

Diese Metriken sind unabhängig von den kostenlosen Metriken, die Amazon EMR unter dem AWS/ElasticMapReduce Namespace in bereitstellt. CloudWatch Weitere Informationen zu den Metriken, die der CloudWatch Agent nicht veröffentlicht, finden Sie unter [Überwachen von Amazon EMR-Metriken mit CloudWatch](#).

Mit Amazon EMR 7.0 und höher können Sie eine benutzerdefinierte Version des CloudWatch Amazon-Agenten auf Ihrem Cluster installieren, um Metriken aus Ihrem EMR-Cluster zu sammeln. Der Agent sammelt die Metriken von allen Knoten im Cluster, sammelt sie auf dem primären Knoten und veröffentlicht die Metriken von diesem Knoten in der Cloud.

Sie können den Agenten installieren, wenn Sie einen neuen Cluster in der Konsole erstellen oder wenn Sie die `create-cluster`-API verwenden. Weitere Informationen finden Sie unter [Erstellen Sie einen EMR-Cluster, der Amazon CloudWatch Agent verwendet](#). Standardmäßig veröffentlicht ein Amazon EMR-Cluster, auf dem die Version 7.x-Serie ausgeführt wird, [34 Metriken auf Systemebene](#) in Intervallen von 60 Sekunden. Sie können den Agenten jedoch so konfigurieren, dass er verschiedene Metriken veröffentlicht. CloudWatch Eine weitere Option besteht darin, Metriken in Amazon Managed Service for Prometheus zu veröffentlichen, und Sie können wählen, welche Metriken ebenfalls veröffentlicht werden sollen. Für verschiedene Anwendungsfälle und

Konfigurationen können Sie den Agenten so konfigurieren, dass er die Metriken in der CloudWatch Konsole, Amazon Managed Grafana oder über die APIs für CloudWatch oder Amazon Managed Service für Prometheus anzeigt und abfragt. Für diese AWS-Services fallen Gebühren an, wenn Sie sie zum Speichern und Abfragen von Metriken verwenden.

Der CloudWatch Agent auf Amazon EMR kann sowohl Systemmetriken als auch JMX-Metriken für die folgenden Services in Ihrem Amazon EMR-Cluster veröffentlichen.

- Hadoop DataNode
- Hadoop NameNode
- Garn NodeManager
- Garn ResourceManager
- HBase Master — nur Amazon EMR 7.1+
- HBase RegionServer — nur Amazon EMR 7.1+
- HBase ThriftServer — nur Amazon EMR 7.1+

Weitere Informationen zu verfügbaren Metriken und zur Konfiguration des CloudWatch Agenten auf Amazon EMR finden Sie unter [CloudWatch Agent für Amazon EMR konfigurieren](#).

In der folgenden Tabelle sind die Version von AmazonCloudWatchAgent aufgeführt, die in der neuesten Version der Amazon EMR 7.x-Serie enthalten ist, zusammen mit den Komponenten, mit denen Amazon EMR installiert wird. AmazonCloudWatchAgent

Informationen zur Version der Komponenten, mit denen AmazonCloudWatchAgent in dieser Version installiert wurde, finden Sie unter [Komponentenversionen der Version 7.1.0](#).

#### AmazonCloudWatchAgent Versionsinformationen für emr-7.1.0

Amazon-EMR-Version sbezeichnung	AmazonCloudWatchAgent Version	Komponenten, die mit installiert wurden AmazonClo udWatchAgent
emr-7.1.0	AmazonCloudWatchAgent 1.300032.2	adot-java-agent, emrfs, emr- amazon-cloudwatch-agent , emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hd

Amazon-EMR-Version sbezeichnung	AmazonCloudWatchAgent Version	Komponenten, die mit installiert wurden AmazonClo udWatchAgent
		fs-datanode, hadoop-hdfs- library, hadoop-hdfs-nameno de, hadoop-httpfs-server, hadoop-kms-server, hadoop- mapred, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server

## Themen

- [Erstellen Sie einen EMR-Cluster, der Amazon CloudWatch Agent verwendet](#)
- [Standardmetriken für CloudWatch Agenten mit Amazon EMR](#)
- [CloudWatch Agent für Amazon EMR konfigurieren](#)
- [Überlegungen und Einschränkungen](#)
- [CloudWatch Versionsverlauf des Agenten](#)

## Erstellen Sie einen EMR-Cluster, der Amazon CloudWatch Agent verwendet

Die Verfahren in diesem Abschnitt beschreiben die Schritte zum Erstellen eines Clusters in Amazon EMR mit dem CloudWatch Amazon-Agenten aus dem AWS Management Console und dem AWS CLI.

## Themen

- [Erforderliche IAM-Berechtigungen für den Agenten CloudWatch](#)
- [Erforderlicher CloudWatch Agenten-Endpunkt](#)
- [Einen EMR-Cluster erstellen](#)

## Erforderliche IAM-Berechtigungen für den Agenten CloudWatch

Der CloudWatch Agent benötigt die AWS Identity and Access Management (IAM-) `cloudwatch:PutMetricData` Berechtigung im Amazon EC2 EC2-Instance-Profil für Amazon EMR. Die Amazon-EMR-Standardrolle verfügt bereits über diese Berechtigung. Sie können die Standardrolle von `with` aus erstellen. AWS CLI `aws emr create-default-roles` Weitere Informationen finden Sie unter [Servicerolle für Cluster-EC2-Instances \(EC2-Instance-Profil\)](#) im Verwaltungshandbuch für Amazon EMR.

Die folgende IAM-Beispielrichtlinie enthält die `cloudwatch:PutMetricData`-Berechtigung:

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": "cloudwatch:PutMetricData",
      "Resource": "*"
    }
  ]
}
```

## Erforderlicher CloudWatch Agenten-Endpunkt

Um Metriken CloudWatch für einen EMR-Cluster in einem privaten Subnetz zu veröffentlichen, erstellen Sie einen CloudWatch Agent-Endpunkt und ordnen Sie ihn der VPC zu, in der sich das private Subnetz befindet.

Weitere Informationen zu den jeweiligen CloudWatch AWS-Region Endpunkten finden Sie unter [CloudWatch Amazon-Endpunkte und Kontingente](#) im AWS Allgemeinen Referenzhandbuch.

## Einen EMR-Cluster erstellen

Nachdem Sie die erforderlichen Berechtigungen und den Endpunkt für die Verwendung mit dem CloudWatch Agenten eingerichtet haben, verwenden Sie den AWS Management Console oder, um einen neuen Cluster AWS CLI zu erstellen, auf dem der Agent installiert ist.

## Console

Um einen Cluster mit Amazon CloudWatch Agent von der Konsole aus zu erstellen

1. Navigieren Sie zur Amazon-EMR-Konsole.
2. Wählen Sie Cluster erstellen.
3. Wählen Sie unter Name and applications (Name und Anwendungen) Amazon-EMR-Version 7.0.0 oder höher aus.
4. Wählen Sie unter Anwendungspaket das Paket oder die Apps aus, die Sie in Ihrem Cluster installieren möchten, und schließen Sie CloudWatch Agenten in Ihre Auswahl ein.
5. Fahren Sie mit der Erstellung des Clusters fort, um den Anforderungen Ihres Anwendungsfalls gerecht zu werden.

## AWS CLI

In der AWS CLI können Sie einen CloudWatch Amazon-Agenten mit dem `--applications` Parameter für zu einem Cluster hinzufügen `create-cluster`.

Um einen Cluster mit CloudWatch Amazon-Agent aus dem AWS CLI

- Wenn Sie einen Cluster erstellen, verwenden Sie einen Befehl, der dem folgenden ähnelt, um den CloudWatch Amazon-Agenten einzubeziehen. Ersetzen Sie *myKey* durch den Namen Ihres EC2-Schlüsselpaars.

```
aws emr create-cluster --name "Spark cluster with CloudWatch agent" \  
--release-label emr-7.0.0 \  
--applications Name=Spark Name=AmazonCloudWatchAgent \  
--ec2-attributes KeyName=myKey --instance-type m7g.2xlarge \  
--instance-count 3 --use-default-roles
```

Weitere Informationen zur Verwendung von Amazon EMR mit dem finden Sie in der AWS CLI [AWS CLI Befehlsreferenz](#).

## Standardmetriken für CloudWatch Agenten mit Amazon EMR

Wenn Sie den CloudWatch Amazon-Agenten auf Amazon EMR installieren, veröffentlicht die Standardkonfiguration die folgenden Systemmetriken für alle Instances in Ihrem Cluster, sofern Sie



[den Agenten nicht anders konfigurieren](#). Definitionen der einzelnen [Metriken finden Sie unter Vom CloudWatch Agenten gesammelte Metriken](#) im [CloudWatch Amazon-Benutzerhandbuch](#).

## CPU

### CPU-Metriken

- `cpu_usage_active`
- `cpu_usage_guest`
- `cpu_usage_guest_nice`
- `cpu_usage_idle`
- `cpu_usage_iowait`
- `cpu_usage_irq`
- `cpu_usage_nice`
- `cpu_usage_softirq`
- `cpu_usage_steal`
- `cpu_usage_system`
- `cpu_usage_user`

## Disk

### Festplattenmetriken

- `disk_free`
- `disk_total`
- `disk_used`
- `disk_used_percent`

## Memory

### Speichermetriken

- `mem_active`
- `mem_available`
- `mem_available_percent`

- mem\_free
- mem\_inactive
- mem\_total
- mem\_used
- mem\_used\_percent
- mem\_buffered
- mem\_cached

## Network IO

### Netzwerk-I/O-Metriken

- net\_bytes\_recv
- net\_bytes\_sent
- net\_packets\_recv
- net\_packets\_sent

## Process

### Metriken verarbeiten

- processes\_running
- processes\_total

## Swap

### Metriken wechseln

- swap\_free
- swap\_used
- swap\_used\_percent

Standardmäßig veröffentlicht der Agent alle Metriken CloudWatch unter dem benutzerdefinierten Namespace CWAgent und unter dem Schemacluster.id,,

`instance.id`, `node.type`, `service.name`. Gehen Sie wie folgt vor, um diese Standardmetriken anzuzeigen, die der CloudWatch Agent für Amazon EMR veröffentlicht:

1. Navigieren Sie zur CloudWatch Amazon-Konsole.
2. Wählen Sie **Metrics** (Metriken) und dann **All metrics** (Alle Metriken) aus.
3. Wählen Sie unter **Custom Namespaces** (Benutzerdefinierte Namespaces) **CWAgent** aus. Wählen Sie dann das Schema aus: **`cluster.id`**, **`instance.id`**, **`node.type`**, **`service.name`**.
4. Fragen Sie die Metriken weiterhin CloudWatch über die Benutzeroberfläche ab, wie es Ihr Anwendungsfall erfordert.

Amazon EMR fügt jeder Metrik die folgenden Bezeichnungen hinzu: `["cluster.id", "instance.id", "node.type", "service.name"]`. Beachten Sie in Bezug auf diese Labels Folgendes:

- Der Wert des `cluster.id`-Labels ist die ID des EMR-Clusters, der die Metrik erzeugt hat. Ein Beispielwert für das `cluster.id`-Label ist: `j-123456789ABC`.
- Der Wert des `instance.id`-Labels ist die ID der Instance in dem EMR-Cluster, der die Metrik erzeugt hat. Ein Beispielwert für das `instance.id`-Label ist `i-01bcf5f140f3355777`.
- Der Wert des `node.type`-Labels steht für den Knotentyp der Instance im Cluster, die die Metrik erzeugt hat. Dieser Wert kann entweder `mastercore`, oder `seintask`.
- Der Wert des `service.name`-Labels ist der Name des Services, der die Metrik erzeugt hat. Der `service.name`-Standardwert für die Standardsystemmetriken ist `system`.

#### Note

Amazon EMR Version 7.0.0 verwendet ein etwas anderes Metrik-Label-Schema. `cluster.id` ist `jobflow.id` und `node.type` ist nicht verfügbar. `instance.id` und `service.name` sind verfügbar.

## CloudWatch Agent für Amazon EMR konfigurieren

Amazon EMR 7.0.0 und höher beinhalten den CloudWatch Amazon-Agenten. In den folgenden Abschnitten wird beschrieben, wie Sie den Agenten auf Ihrem EMR-Cluster konfigurieren, wie unter [beschrieben Erstellen Sie einen EMR-Cluster, der Amazon CloudWatch Agent verwendet](#).

## Themen

- [CloudWatch Agent für Amazon EMR 7.1.0 konfigurieren](#)
- [CloudWatch Agent für Amazon EMR 7.0.0 konfigurieren](#)

## CloudWatch Agent für Amazon EMR 7.1.0 konfigurieren

Ab Amazon EMR 7.1.0 können Sie den CloudWatch Amazon-Agenten so konfigurieren, dass er zusätzliche Systemmetriken verwendet, Anwendungsmetriken hinzufügt und das Metrikziel mithilfe der Amazon EMR-Konfigurations-API ändert. Weitere Informationen zur Verwendung der EMR-Konfigurations-API zur Konfiguration der Clusteranwendungen finden [Sie unter Anwendungen konfigurieren](#).

### Note

7.1.0 unterstützt nur den Rekonfigurationstyp. OVERWRITE Weitere Informationen zu den Rekonfigurationstypen finden [Sie unter Überlegungen bei der Neukonfiguration](#) einer Instanzgruppe.

## Themen

- [Konfigurationsschema](#)
- [Beispiele für Konfigurationen von Systemmetriken](#)
- [Beispiele für Konfigurationen von Anwendungsmetriken](#)
- [Beispiel für Amazon Managed Service für Prometheus](#)

## Konfigurationsschema

`emr-metricshat` die folgenden Klassifizierungen:

- `emr-system-metrics`— konfiguriert Systemmetriken wie CPU, Festplatte und Speicher.
- `emr-hadoop-hdfs-datanode-metrics`— Hadoop DataNode JMX-Metriken konfigurieren
- `emr-hadoop-hdfs-namenode-metrics`— Hadoop JMX-Metriken NameNode konfigurieren
- `emr-hadoop-yarn-nodemanager-metrics`— Yarn JMX-Metriken NodeManager konfigurieren
- `emr-hadoop-yarn-resourcemanager-metrics`— Yarn ResourceManager JMX-Metriken konfigurieren

- `emr-hbase-master-metrics`— HBase Master JMX-Metriken konfigurieren
- `emr-hbase-region-server-metrics`— JMX-Metriken für HBase Region Server konfigurieren
- `emr-hbase-rest-server-metrics`— HBase REST Server-JMX-Metriken konfigurieren
- `emr-hbase-thrift-server-metrics`— HBase Thrift Server JMX-Metriken konfigurieren

In den folgenden Tabellen werden die verfügbaren Eigenschaften und Konfigurationen für alle Klassifizierungen beschrieben.

### Eigenschaften von emr-Metrics

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	Hinweise
<code>metrics_destination</code>	Optional	Legt fest, ob Cluster-Metriken auf Amazon CloudWatch oder Amazon Managed Service for Prometheus veröffentlicht werden.	„CLOUDWATCH“	„CLOUDWATCH“, „PROMETHEUS“	Bei dieser Eigenschaft wird nicht zwischen Groß- und Kleinschreibung unterschieden. Zum Beispiel ist „Cloudwatch“ dasselbe wie „CLOUDWATCH“.
<code>prometheus_endpoint</code>	Optional	Wenn auf „PROMETHEUS“ gesetzt <code>metrics_destination</code> ist, konfiguriert diese	N/A	Jede gültige Remote-Write-URL für Amazon Managed Service für Prometheus. Das URL-	Dieses Feld ist erforderlich, wenn <code>metrics_destination</code> es auf „PROMETHEUS“ gesetzt

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	Hinweise
		Eigenschaft den CloudWatch Agenten so, dass er Metriken an den bereitgestellten Remote-Write-Endpoint von Amazon Managed Service for Prometheus sendet.		Format für das Remote-Schreiben ist  <pre>https:// aps- worksp aces. &lt;region&gt; s.com/ wor kspaces/ &lt;workspace_id&gt; / api/v1/r emote_wri te</pre>	ist. Die Bereitstellung schlägt fehl, wenn Sie keinen Schlüssel angeben oder wenn der Wert eine leere Zeichenfolge ist.

### emr-system-metrics Eigenschaften

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	Hinweise
<code>metrics_collection_interval</code>	Optional	Wie oft in Sekunden Metriken vom CloudWatch Agenten gesammelt und veröffentlicht werden.	„60“	Eine Zeichenfolge, die die Anzahl der Sekunden angibt. Akzeptiert nur ganze Zahlen.	Sie können diese Eigenschaft mit der <code>metrics_collection_interval</code> Eigenschaft einzelner Metrikgruppen

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	Hinweise
-------------	--------------	--------------	--------------	----------------	----------

überschreiben.

## emr-system-metrics Konfigurationen

### cpu

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	Hinweise
<code>metrics</code>	Optional	Die Liste der CPU-Metriken, die der Agent sammeln soll.	Siehe <a href="#">Standardmetriken für CloudWatch Agenten mit Amazon EMR</a>	Eine durch Kommas getrennte Liste gültiger CPU-Metrikenamen mit oder ohne <code>cpu_</code> Präfix, z. B. und <code>usage_active</code> <code>cpu_time_idle</code> Gültige <a href="#">Metriken finden Sie unter Vom CloudWatch Agenten gesammelte Metriken.</a>	Die Angabe einer leeren Zeichenfolge bedeutet, dass keine CPU-Metriken veröffentlicht werden.
<code>metrics_collection</code>	Optional	Wie oft (in Sekunden)	Der Wert des <code>globalmetrics</code>	Eine Zeichenfolge	Dieser Wert überschreibt

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	Hinweise
<code>_interval</code>		der Agent CPU-Metriken sammeln und veröffentlichen soll.	<code>collection_interval</code>	Die Anzahl der Sekunden, die die Agent CPU-Metriken sammeln und veröffentlichen soll. Akzeptiert nur ganze Zahlen.	die globale <code>metrics_collection_interval</code> Eigenschaft nur für CPU-Metriken.
<code>drop_original_metrics</code>	Optional	Liste der CPU-Metriken, für die keine unaggregierten Metriken veröffentlicht werden sollen.	Es wurden keine unaggregierten CPU-Metriken veröffentlicht.	Eine durch Kommas getrennte Liste von CPU-Metriken, die auch in der Metrik-Eigenschaft angegeben sind. Eine leere Zeichenfolge bedeutet, dass alle CPU-Metriken veröffentlicht werden.	Der CloudWatch Agent aggregiert alle Metriken nach Cluster-ID, Instanz-ID, Knotentyp und Dienstname. Standardmäßig veröffentlicht der CloudWatch Agent keine ressourcenspezifischen Metriken für Metriken mit mehreren Ressourcen.



Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	Hinweise
<code>resources</code>	Optional	Legt fest, ob der Agent Metriken pro Kern veröffentlicht.	"*"	„*“ aktiviert Metriken pro Kern. „“ deaktiviert Metriken pro Kern.	Der CloudWatch Agent veröffentlicht nur Metriken pro Kern für CPU-Metriken, die nicht hinzugefügt wurden. <code>drop_original_metrics</code>

## disk

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	Hinweise
<code>metrics</code>	Optional	Die Liste der Festplatten-Metriken, die der Agent sammeln soll.	Siehe <a href="#">Standardmetriken für CloudWatch Agenten mit Amazon EMR</a>	Eine durch Kommas getrennte Liste gültiger Festplatten-Metrikenamen mit oder ohne <code>disk_</code> -Präfix, z. B. und <code>disk_totalused_percent</code> Gültige	Die Angabe einer leeren Zeichenfolge bedeutet, dass keine Festplattenmetriken veröffentlicht werden.

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	Hinweise
				<a href="#">Messwerte finden Sie unter <u>Vom CloudWatch Agenten gesammelte Metriken.</u></a>	
<code>metrics_collection_interval</code>	Optional	Gibt an, wie oft in Sekunden der Agent Festplattenmetriken sammeln und veröffentlichen soll.	Der Wert des <code>globalmetrics_collection_interval</code> .	Eine Zeichenfolge, die die Anzahl der Sekunden angibt. Akzeptiert nur ganze Zahlen.	Dieser Wert überschreibt die globale <code>metrics_collection_interval</code> Eigenschaft nur für Festplattenmesswerte.

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	Hinweise
<code>drop_original_metrics</code>	Optional	Liste der Festplattenmesswerte, für die keine unaggregierten Messwerte veröffentlicht werden sollen.	Es wurden keine unaggregierten Festplattenmetriken veröffentlicht.	Eine durch Kommas getrennte Liste von Festplattenmetriken, die auch in der Metrik-Eigenschaft angegeben sind. Eine leere Zeichenfolge bedeutet, dass alle Festplattenmetriken veröffentlicht werden.	Der CloudWatch Agent aggregiert alle Metriken nach Cluster-ID, Instanz-ID, Knotentyp und Dienstname. Standardmäßig veröffentlicht der CloudWatch Agent keine ressourcenspezifischen Metriken für Metriken mit mehreren Ressourcen.

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	Hinweise
<code>resources</code>	Optional	Legt fest, ob der Agent per-mount-point Metriken veröffentlicht.	"*"	„*“ bedeutet alle Einhängpunkte, „“ bedeutet keine Einhängpunkte oder eine durch Kommas getrennte Liste von Bereitstellungs-punkten. z. B. <code>"/, /emr"</code> .	Der CloudWatch Agent veröffentlicht nur per-mount-point Metriken für Festplattenmetriken, die nicht übernommen wurden. <code>drop_original_metrics</code>

## diskio

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	Hinweise
<code>metrics</code>	Optional	Die Liste der Festplatten-E/A-Metriken, die der Agent sammeln soll.	Siehe <a href="#">Standardmetriken für CloudWatch Agenten mit Amazon EMR</a>	Eine durch Kommas getrennte Liste gültiger Festplatten-IO-Metrikenamen mit oder ohne <code>diskio_</code> Präfix, z. B. und.	Die Angabe einer leeren Zeichenfolge bedeutet, dass keine Festplatten-IO-Metriken veröffentlicht werden.

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	Hinweise
				diskio_reads writes Gültige <a href="#">Metriken</a> <a href="#">finden Sie unter Vom CloudWatch Agenten gesammelte Metriken.</a>	
metrics_collection_interval	Optional	Gibt an, wie oft in Sekunden der Agent Festplatten-E/A-Metriken sammeln und veröffentlichen soll.	Der Wert des globalen metrics_collection_interval .	Eine Zeichenfolge, die die Anzahl der Sekunden angibt. Akzeptiert nur ganze Zahlen.	Dieser Wert überschreibt die globale metrics_collection_interval Eigenschaft nur für Festplatten-E/A-Metriken.

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	Hinweise
<code>drop_original_metrics</code>	Optional	Liste der Festplatten-IO-Metriken, für die keine unaggregierten Metriken veröffentlicht werden sollen.	Es wurden keine unaggregierten Festplatten-E/A-Metriken veröffentlicht.	Eine durch Kommas getrennte Liste von Festplatten-E/A-Metriken, die auch in der Metrik-Eigenschaft angegeben sind. Eine leere Zeichenfolge bedeutet, dass alle Festplatten-IO-Metriken veröffentlicht werden.	Der CloudWatch Agent aggregiert alle Metriken nach Cluster-ID, Instanz-ID, Knotentyp und Dienstname. Standardmäßig veröffentlicht der CloudWatch Agent keine ressourcenspezifischen Metriken für Metriken mit mehreren Ressourcen.

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	Hinweise
<code>resources</code>	Optional	Legt fest, ob der Agent Metriken pro Gerät veröffentlicht.	<code>"*"</code>	„*“ bedeutet alle Speichergereäte, „“ bedeutet keine Speichergereäte oder eine durch Kommas getrennte Liste von Gerätenamen. z. B. <code>"nvme0n1, nvme1n1"</code> .	Der CloudWatch Agent veröffentlicht nur Metriken pro Gerät für Festplatten-E/A-Metriken, die nicht übernommen wurden. <code>drop_original_metrics</code>

## mem

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	Hinweise
<code>metrics</code>	Optional	Die Liste der Speichermetriken, die der Agent sammeln soll.	Siehe <a href="#">Standardmetriken für CloudWatch Agenten mit Amazon EMR</a>	Eine durch Kommas getrennte Liste gültiger Speichermetriken mit oder ohne <code>mem_</code> Präfix, z. B. und <code>mem_avail</code>	Die Angabe einer leeren Zeichenfolge bedeutet, dass keine Speichermetriken veröffentlicht werden.

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	Hinweise
				able available _percent Gültige <a href="#">Metriken finden Sie unter Vom CloudWatch Agenten gesammelte Metriken.</a>	
metrics_collection_interval	Optional	Gibt an, wie oft in Sekunden der Agent Speichermetriken sammeln und veröffentlichen soll.	Der Wert des globalenmetrics_collection_interval.	Eine Zeichenfolge, die die Anzahl der Sekunden angibt. Akzeptiert nur ganze Zahlen.	Dieser Wert überschreibt die globale metrics_collection_interval Eigenschaft nur für Speichermetriken.

## net

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	Hinweise
metrics	Optional	Die Liste der Netzwerkmetriken, die der Agent	Siehe <a href="#">Standardmetriken für CloudWatch Agenten</a>	Eine durch Kommas getrennte Liste gültiger Netzwerk	Die Angabe einer leeren Zeichenfolge bedeutet, dass keine



Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	Hinweise
		sammeln soll.	<a href="#">mit Amazon EMR</a>	etrikenname mit oder ohne net_ Präfix, z. B. und. net_packets_sent packets_received Gültige <a href="#">Metriken finden Sie unter Vom CloudWatch Agenten gesammelte Metriken.</a>	Netzwerkmetriken veröffentlicht werden.
metrics_collection_interval	Optional	Wie oft in Sekunden sollte der Agent Netzwerkmetriken sammeln und veröffentlichen.	Der Wert des Globalenmetrics_collection_interval .	Eine Zeichenfolge, die die Anzahl der Sekunden angibt. Akzeptiert nur ganze Zahlen.	Dieser Wert überschreibt die globale metrics_collection_interval Eigenschaft nur für Netzwerkmetriken.

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	Hinweise
<code>drop_original_metrics</code>	Optional	Liste der Netzwerkmetriken, für die keine unaggregierten Messwerte veröffentlicht werden sollen.	Es wurden keine unaggregierten Netzwerkmetriken veröffentlicht.	Eine durch Kommas getrennte Liste von Netzwerkmetriken, die auch in der Metrik-Eigenschaft angegeben sind. Eine leere Zeichenfolge bedeutet, dass alle Netzwerkmetriken veröffentlicht werden.	Der CloudWatch Agent aggregiert alle Metriken nach Cluster-ID, Instanz-ID, Knotentyp und Dienstname. Standardmäßig veröffentlicht der CloudWatch Agent keine ressourcenspezifischen Metriken für Metriken mit mehreren Ressourcen.

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	Hinweise
<code>resources</code>	Optional	Legt fest, ob der Agent Metriken pro Schnittstelle veröffentlicht.	"*"	„*“ bedeutet alle Netzwerkschnittstellen, „“ bedeutet keine Netzwerkschnittstellen oder eine durch Kommas getrennte Liste von Schnittstellennamen. z. B. "eth0,eth1" .	Der CloudWatch Agent veröffentlicht nur Schnittstellen-Metriken für Netzwerkmetriken, die nicht hinzugefügt wurden. <code>drop_original_metrics</code>

## netstat

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	Hinweise
<code>metrics</code>	Optional	Die Liste der Netzwerkstatistik-Metriken, die der Agent sammeln soll.	Siehe <a href="#">Standardmetriken für CloudWatch Agenten mit Amazon EMR</a>	Eine durch Kommas getrennte Liste gültiger Speichermetriken mit oder ohne <code>netstat_</code>	Die Angabe einer leeren Zeichenfolge bedeutet, dass keine Netzwerkstatistik-Metriken

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	Hinweise
				Präfix, z. B. und. tcp_listen netstat_udp_socket Gültige <a href="#">Metriken finden Sie unter <u>Vom CloudWatch Agenten gesammelte Metriken.</u></a>	veröffentlicht werden.
metrics_collection_interval	Optional	Wie oft in Sekunden sollte der Agent Netzwerkstatistiken sammeln und veröffentlichen.	Der Wert des metrics_collection_interval Globalen.	Eine Zeichenfolge, die die Anzahl der Sekunden angibt. Akzeptiert nur ganze Zahlen.	Dieser Wert überschreibt die globale metrics_collection_interval Eigenschaft nur für Netzwerkstatistiken.

## processes

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	Hinweise
<code>metrics</code>	Optional	Die Liste der Prozessmetriken, die der Agent sammeln soll.	Siehe <a href="#">Standardmetriken für CloudWatch Agenten mit Amazon EMR</a>	Eine durch Kommas getrennte Liste gültiger Speichermetriken mit oder ohne <code>processes_</code> Präfix, z. B. und <code>processes_running_total</code> . Gültige <a href="#">Metriken</a> finden Sie unter <a href="#">Vom CloudWatch Agenten gesammelte Metriken</a> .	Die Angabe einer leeren Zeichenfolge bedeutet, dass keine Prozessmetriken veröffentlicht werden.
<code>metrics_collection_interval</code>	Optional	Wie oft in Sekunden sollte der Agent Systemprozessmetriken sammeln und veröffentlichen.	Der Wert des <code>Globalenmetrics_collection_interval</code> .	Eine Zeichenfolge, die die Anzahl der Sekunden angibt. Akzeptiert nur ganze Zahlen.	Dieser Wert überschreibt die globale <code>metrics_collection_interval</code> Eigenschaft nur für Systempro

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	Hinweise
					zessmetriken.

## swap

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	Hinweise
<code>metrics</code>	Optional	Die Liste der Swap-Metriken, die der Agent sammeln soll.	Siehe <a href="#">Standardmetriken für CloudWatch Agenten mit Amazon EMR</a>	Eine durch Kommas getrennte Liste gültiger Speichermetriken mit oder ohne <code>swap_</code> Präfix, z. B. und <code>swap_free</code>  <code>used_percent</code> Gültige <a href="#">Metriken finden Sie unter Vom CloudWatch Agenten gesammelte Metriken.</a>	Die Angabe einer leeren Zeichenfolge bedeutet, dass keine Swap-Metriken veröffentlicht werden.
<code>metrics_collection</code>	Optional	Wie oft in Sekunden sollte der Agent Swap-	Der Wert des <code>Globalmetrics_collection</code>	Eine Zeichenfolge, die die Anzahl der	Dieser Wert überschreibt die globale

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	Hinweise
<code>_interval</code>		Metriken sammeln und veröffentlichen.	<code>_interval</code> .	Sekundenangibt. Akzeptiert nur ganze Zahlen.	<code>metrics_collection_interval</code> Eigenschaft nur für Swap-Metriken.

### emr-hadoop-hdfs-datanodeEigenschaften von -metrics

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	
<code>&lt;custom_bean_name&gt;</code>	Optional	N/A	Die MBean, von der der CloudWatch Agent Metriken sammeln soll, z. B. <code>Hadoop:service=DataNode,name=DataNodeActivity</code> Beispiel-MBean-Namen und die entsprechenden Metriken finden Sie in den <a href="#">JMX-</a>	Eine Zeichenfolge, die die kommagetrennte Liste der Metriken enthält, die mit der MBean verknüpft sind. z. B. <code>BlocksCached,BlockRead</code> .	

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	
			<a href="#">YAML-Beispielateien</a> für Amazon EMR Version 7.0.		
<code>otel.metric.export.interval</code>	Optional	Wie oft in Millisekunden Hadoop-Metriken erfasst werden sollen. DataNode	„60000“	Eine Zeichenfolge, die die Anzahl der Millisekunden angibt. Akzeptiert nur ganze Zahlen.	

## emr-hadoop-hdfs-namenodeEigenschaften von -metrics

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	
<code>&lt;custom_bean_name&gt;</code>	Optional	N/A	Die MBean, von der der CloudWatch Agent Metriken sammeln soll, z. B. <code>Hadoop:service=NameNode,name=FSNamesystem</code> Beispiel-	Eine Zeichenfolge, die die kommagetrennte Liste der Metriken enthält, die mit der MBean verknüpft sind. z. B. <code>BlockCapacity, Capa</code>	



Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	
			MBean-Namen und die entsprechenden Metriken finden Sie in den <a href="#">JMX-YAML-Beispielen</a> für Amazon EMR Version 7.0.	cityUsedGB .	
<code>otel.metric.export.interval</code>	Optional	Wie oft in Millisekunden Hadoop-Metriken erfasst werden sollen. NameNode	„60000“	Eine Zeichenfolge, die die Anzahl der Millisekunden angibt. Akzeptiert nur ganze Zahlen.	

#### emr-hadoop-yarn-nodemanagerEigenschaften von -metrics

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	
<code>&lt;custom_bean_name&gt;</code>	Optional	N/A	Die MBean, von der der CloudWatch Agent Metriken sammeln	Eine Zeichenfolge, die die kommagetrennte Liste der Metriken	

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte
			soll, z. B. Hadoop:service=NodeManager,name=NodeManagerMetrics Beispiel-MBean-Namen und die entsprechenden Metriken finden Sie in den <a href="#">JMX-YAML-Beispielen</a> für Amazon EMR Version 7.0.	enthält, die mit der MBean verknüpft sind. z. B. MaxCapacity, AllocatedGB .
otel.metric.export.interval	Optional	Wie oft in Millisekunden Hadoop YARN-Metriken erfasst werden sollen. NodeManager	„60000“	Eine Zeichenfolge, die die Anzahl der Millisekunden angibt. Akzeptiert nur ganze Zahlen.

emr-hadoop-yarn-resourcemanagerEigenschaften von -metrics

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte
<code>&lt;custom_bean_name&gt;</code>	Optional	N/A	Die MBean, von der der CloudWatch Agent Metriken sammeln soll, z. B. <code>Hadoop:service=ResourceManager,name=PartitionQueueMetrics</code> . Beispiel-MBean-Namen und die entsprechenden Metriken finden Sie in den <a href="#">JMX-YAML-Beispielen</a> für Amazon EMR Version 7.0.	Eine Zeichenfolge, die die kommagetrennte Liste der Metriken enthält, die mit der MBean verknüpft sind. z. B. <code>MaxCapacity,MaxCapacityVCores</code> .
<code>otel.metric.export.interval</code>	Optional	Wie oft in Millisekunden Hadoop YARN-Metriken erfasst	„60000“	Eine Zeichenfolge, die die Anzahl der Millisekunden

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	
		werden sollen. ResourceManager		nden angibt. Akzeptiert nur ganze Zahlen.	

### emr-hbase-master-metrics Eigenschaften

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	
<code>&lt;custom_bean_name&gt;</code>	Optional	N/A	Die MBean, von der der CloudWatch Agent Metriken sammeln soll, z. B. <code>Hadoop:service=HBase,name=Master,sub=AssignmentManager</code> Beispiel-MBean-Namen und die entsprechenden Metriken finden Sie in den <a href="#">JMX-YAML-Beispiel</a>	Eine Zeichenfolge, die die kommagetrennte Liste der Metriken enthält, die mit der MBean verknüpft sind. z. B. <code>AssignFailedCount, AssignSubmittedCount</code> .	

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	
			<a href="#">ateien</a> für Amazon EMR Version 7.0.		
<code>otel.metric.export.interval</code>	Optional	Wie oft in Millisekunden HBase Master-Metriken erfasst werden sollen.	„60000“	Eine Zeichenfolge, die die Anzahl der Millisekunden angibt. Akzeptiert nur ganze Zahlen.	

#### emr-hbase-region-serverEigenschaften von -metrics

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	
<code>&lt;custom_bean_name&gt;</code>	Optional	N/A	Die MBean, von der der CloudWatch Agent Metriken sammeln soll, z. B. <code>Hadoop:service=HBase,name=RegionServer,sub=IPC</code> Beispiel-MBean-Nam	Eine Zeichenfolge, die die kommagetrennte Liste der Metriken enthält, die mit der MBean verknüpft sind. z. B. <code>numActiveHandler,numActiveP</code>	

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	
			en und die entsprechenden Metriken finden Sie in den <a href="#">JMX-YAML-Beispielen</a> für Amazon EMR Version 7.0.	riorityHandler .	
otel.metric.export.interval	Optional	Wie oft in Millisekunden HBase Region Server-Metriken erfasst werden sollen.	„60000“	Eine Zeichenfolge, die die Anzahl der Millisekunden angibt. Akzeptiert nur ganze Zahlen.	

#### emr-hbase-rest-serverEigenschaften von -metrics

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	
<i>&lt;custom_bean_name&gt;</i>	Optional	N/A	Die MBean, von der der CloudWatch Agent Metriken sammeln soll, z. B.	Eine Zeichenfolge, die die kommagetrennte Liste der Metriken enthält,	

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte
			Hadoop:service=HBase,name=REST Beispiel-MBean-Namen und die entsprechenden Metriken finden Sie in den <a href="#">JMX-YAML-Beispielen</a> für Amazon EMR Version 7.0.	die mit der MBean verknüpft sind. z. B. successfulPut, successfulScanCount .
otel.metric.export.interval	Optional	Wie oft in Millisekunden HBase Rest Server-Metriken erfasst werden sollen.	„60000“	Eine Zeichenfolge, die die Anzahl der Millisekunden angibt. Akzeptiert nur ganze Zahlen.

### emr-hbase-thrift-serverEigenschaften von -metrics

Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte
<code>&lt;custom_bean_name&gt;</code>	Optional	N/A	Die MBean, von der der CloudWatch Agent Metriken sammeln soll, z. B. <code>Hadoop:service=HBase,name=Thrift,sub=ThriftOn</code> . e Beispiel-MBean-Namen und die entsprechenden Metriken finden Sie in den <a href="#">JMX-YAML-Beispielen</a> für Amazon EMR Version 7.0.	Eine Zeichenfolge, die die kommagetrennte Liste der Metriken enthält, die mit der MBean verknüpft sind. z. B. <code>BatchGet_max,BatchGet_mean</code> .
<code>otel.metric.export.interval</code>	Optional	Wie oft in Millisekunden sollen HBase Thrift-Servermetriken erfasst werden.	„60000“	Eine Zeichenfolge, die die Anzahl der Millisekunden angibt. Akzeptiert



Eigenschaft	Erforderlich	Beschreibung	Standardwert	Mögliche Werte	
				nur ganze Zahlen.	

## Beispiele für Konfigurationen von Systemmetriken

Das folgende Beispiel zeigt, wie der CloudWatch Agent so konfiguriert wird, dass der Export aller Systemmetriken beendet wird.

```
[
  {
    "Classification": "emr-metrics",
    "Properties": {},
    "Configurations": [
      {
        "Classification": "emr-system-metrics",
        "Properties": {},
        "Configurations": []
      }
    ]
  }
]
```

Im folgenden Beispiel wird der CloudWatch Agent so konfiguriert, dass er die Standard-Systemmetriken exportiert. Auf diese Weise können Sie den Agenten schnell wieder auf den Export der Standard-Systemmetriken zurücksetzen, wenn Sie die Systemmetriken bereits mindestens einmal neu konfiguriert haben. Durch diesen Reset werden auch alle Anwendungsmetriken entfernt, die zuvor neu konfiguriert wurden.

```
[
  {
    "Classification": "emr-metrics",
    "Properties": {},
    "Configurations": []
  }
]
```

Im folgenden Beispiel wird der Cluster so konfiguriert, dass er die `cpumem`, und die `disk` Metriken exportiert.

```
[
  {
    "Classification": "emr-metrics",
    "Properties": {},
    "Configurations": [
      {
        "Classification": "emr-system-metrics",
        "Properties": {
          "metrics_collection_interval": "20"
        },
        "Configurations": [
          {
            "Classification": "cpu",
            "Properties": {
              "metrics": "cpu_usage_guest,cpu_usage_idle",
              "metrics_collection_interval": "30",
              "drop_original_metrics": "cpu_usage_guest"
            }
          },
          {
            "Classification": "mem",
            "Properties": {
              "metrics": "mem_active"
            }
          },
          {
            "Classification": "disk",
            "Properties": {
              "metrics": "disk_used_percent",
              "resources": "/,/mnt",
              "drop_original_metrics": ""
            }
          }
        ]
      }
    ]
  }
]
```

Die vorherige Beispielkonfiguration hat die folgenden Eigenschaften:

- Alle 30 Sekunden erfasst der Agent die `cpu_guest` Metrik für alle CPUs. Sie finden die aggregierte Metrik unter dem CloudWatch Namespace. `CWAgent > cluster.id, instance.id, node.type, service.name`
- Alle 30 Sekunden erfasst der Agent die `cpu_idle` Metrik für alle CPUs. Sie finden die aggregierte Metrik unter dem CloudWatch Namespace. `CWAgent > cluster.id, instance.id, node.type, service.name` Der Agent sammelt auch die Metriken pro CPU. Sie finden sie im selben Namespace. Der Agent sammelt diese Metrik, weil die `drop_original_metrics` Eigenschaft nichts enthält `cpu_idle`, sodass der Agent die Metrik nicht ignoriert.
- Alle 20 Sekunden erfasst der Agent die `mem_active` Metrik. Sie finden die aggregierte Metrik unter dem CloudWatch `CWAgent > cluster.id, instance.id, node.type, service.name` Namespace.
- Alle 20 Sekunden erfasst der Agent die `disk_used_percent` Metriken für die `/` und die `/mnt` Festplatten-Mounts. Sie finden die aggregierten Metriken unter dem CloudWatch Namespace. `CWAgent > cluster.id, instance.id, node.type, service.name` Der Agent sammelt auch die Metriken pro Mount. Sie finden sie im selben Namespace. Der Agent sammelt diese Metrik, weil die `drop_original_metrics` Eigenschaft nichts enthält `disk_used_percent`, sodass der Agent die Metrik nicht ignoriert.

## Beispiele für Konfigurationen von Anwendungsmetriken

Im folgenden Beispiel wird der CloudWatch Agent so konfiguriert, dass der Export von Metriken für den Hadoop Namenode-Dienst beendet wird.

```
[
  {
    "Classification": "emr-metrics",
    "Properties": {},
    "Configurations": [
      {
        "Classification": "emr-hadoop-hdfs-namenode-metrics",
        "Properties": {},
        "Configurations": []
      }
    ]
  }
]
```

Im folgenden Beispiel wird ein Cluster für den Export von Hadoop-Anwendungsmetriken konfiguriert.

```
[
  {
    "Classification": "emr-metrics",
    "Properties": {},
    "Configurations": [
      {
        "Classification": "emr-hadoop-hdfs-namenode-metrics",
        "Properties": {
          "Hadoop:service=NameNode,name=FSNamesystem": "BlockCapacity,CapacityUsedGB",
          "otel.metric.export.interval": "20000"
        },
        "Configurations": []
      },
      {
        "Classification": "emr-hadoop-hdfs-datanode-metrics",
        "Properties": {
          "Hadoop:service=DataNode,name=JvmMetrics": "MemNonHeapUsedM",
          "otel.metric.export.interval": "30000"
        },
        "Configurations": []
      },
      {
        "Classification": "emr-hadoop-yarn-resource-manager-metrics",
        "Properties": {
          "Hadoop:service=ResourceManager,name=CapacitySchedulerMetrics":
"AllocateNumOps,NodeUpdateNumOps"
        },
        "Configurations": []
      }
    ]
  }
]
```

Das vorherige Beispiel hat die folgenden Eigenschaften:

- Alle 20 Sekunden sammelt der Agent die BlockCapacity und CapacityUsedGB -Metriken von Instanzen, auf denen der Hadoop Namenode-Dienst ausgeführt wird.
- Alle 30 Sekunden sammelt der Agent MemNonHeapUsedM Metriken von Instanzen, auf denen der Hadoop Datanode-Dienst ausgeführt wird.
- Alle 30 Sekunden sammelt der Agent die NodeUpdateNumOps Metriken AllocateNumOps und die Metriken von Instanzen, auf denen Hadoop YARN ausgeführt wird. ResourceMaanager

## Beispiel für Amazon Managed Service für Prometheus

Das folgende Beispiel zeigt, wie der CloudWatch Agent so konfiguriert wird, dass er Metriken nach Amazon Managed Service for Prometheus exportiert.

Wenn Sie derzeit Metriken nach Amazon Managed Service for Prometheus exportieren und die Metriken für den Cluster neu konfigurieren und weiterhin Metriken nach Amazon Managed Service for Prometheus exportieren möchten, müssen Sie die Eigenschaften und angeben.

```
metrics_destination prometheus_endpoint
```

```
[
  {
    "Classification": "emr-metrics",
    "Properties": {
      "metrics_destination": "prometheus",
      "prometheus_endpoint": "http://amp-workspace/api/v1/remote_write"
    },
    "Configurations": []
  }
]
```

Verwenden Sie das folgende Beispiel, um den CloudWatch Agenten zum Exportieren von Metriken zu CloudWatch verwenden.

```
[
  {
    "Classification": "emr-metrics",
    "Properties": {
      "metrics_destination": "cloudwatch"
    },
    "Configurations": []
  }
]
```

### Note

Der CloudWatch Agent verfügt über einen Prometheus-Exporter, der bestimmte Attribute umbenennt. Für die Standardkennzahlen verwendet Amazon Managed Service for Prometheus Unterstriche anstelle der Punkte, die Amazon verwendet. CloudWatch Wenn Sie Amazon Managed Grafana verwenden, um die Standardmetriken in Amazon

Managed Service for Prometheus zu visualisieren, werden die Bezeichnungen `alscluster_id`, `instance_id`, `node_type` und angezeigt. `service_name`

## CloudWatch Agent für Amazon EMR 7.0.0 konfigurieren

Sie können den CloudWatch Amazon-Agenten so konfigurieren, dass er zusätzliche Systemmetriken verwendet, die über [die standardmäßige CloudWatch Agentenkonfiguration](#) hinausgehen. Die Konfiguration für 7.0.0 erfordert die Verwendung von Bootstrap-Aktionen, für die wir in den folgenden Abschnitten Beispiele bereitgestellt haben. In einer kommenden Version wird Amazon EMR zusätzliche Konfigurationsoptionen über die Amazon-EMR-API bereitstellen.

### Themen

- [Zusätzliche Systemmetriken mit Amazon EMR 7.0.0 konfigurieren](#)
- [Anwendungsmetriken mit Amazon EMR 7.0.0 konfigurieren](#)
- [Amazon Managed Service für Prometheus als Cloud-Speicher für Metriken mit Amazon EMR 7.0.0 konfigurieren](#)

## Zusätzliche Systemmetriken mit Amazon EMR 7.0.0 konfigurieren

Gehen Sie wie folgt vor, um den Agenten so zu konfigurieren, dass er einen anderen Satz von Systemmetriken in Amazon EMR 7.0.0 verwendet:

1. Erstellen oder wählen Sie einen Bucket in Ihrem Amazon S3 S3-Konto aus, in dem Sie die Konfigurationsdateien speichern möchten, die die CloudWatch Agent-Metriken spezifizieren.
2. Erstellen Sie die `emr-amazon-cloudwatch-agent.json`-Konfigurationsdatei mit den angegebenen bevorzugten Metriken. Verwenden Sie dazu eine der Methoden, die unter [CloudWatch Agenten-Konfigurationsdatei erstellen](#) beschrieben werden. Weitere Informationen zur Struktur der CloudWatch Agenten-Konfigurationsdatei finden Sie unter [Manuelles Erstellen oder Bearbeiten der CloudWatch Agenten-Konfigurationsdatei](#) im CloudWatch Amazon-Benutzerhandbuch.
3. Navigieren Sie als Nächstes zum `aws-emr-utilities` Repo auf GitHub und laden Sie die folgenden Skripte für Systemmetriken herunter:
  - [install\\_system\\_metrics\\_launcher.sh](#) – Ein Skript, das heruntergeladen und dann `install_system_metrics.sh` im Hintergrund ausführt, sodass der Knoten das Bootstrapping beenden kann.

- [install\\_system\\_metrics.sh](#) – Ein Skript, das darauf wartet, dass die Instance, auf der es ausgeführt wird, das Bootstrapping abgeschlossen hat, dann die Konfiguration in der JSON-Datei herunterlädt und anwendet.
4. Öffnen Sie jede SH-Datei und ersetzen Sie *my-s3-bucket* durch den Namen Ihres Buckets aus Schritt 1.
  5. Laden Sie die eine JSON-Datei und die beiden SH-Dateien in Ihren S3-Bucket hoch.
  6. Jetzt können Sie zur Amazon EMR-Konsole navigieren und mit dem CloudWatch Agenten einen neuen Cluster erstellen. Wählen Sie im Navigationsbereich unter EMR in EC2 die Option Cluster und dann Create cluster (Cluster erstellen) aus.
  7. Wählen Sie im Abschnitt Name and applications (Name und Anwendungen) Amazon-EMR-Version 7.0.0 oder höher aus.
  8. Wählen Sie unter Anwendungspaket das Paket oder die Apps aus, die Sie in Ihrem Cluster installieren möchten, und schließen Sie Amazon CloudWatch Agent in Ihre Auswahl ein.
  9. Wählen Sie im Abschnitt Bootstrap actions (Bootstrap-Aktionen) die Option Add (Hinzufügen) aus.
    - Geben Sie für Name **install\_system\_metrics\_launcher.sh** ein.
    - Fügen Sie für Script location (Skript-Standort) **s3://my-s3-bucket/install\_system\_metrics\_launcher.sh** ein. Ersetzen Sie *my-s3-bucket* durch den Pfad zu Ihrem S3-Bucket.
    - Lassen Sie den Argument-Block leer.
  10. Wählen Sie Bootstrap-Aktion hinzufügen aus.
  11. Fahren Sie mit der Erstellung des Clusters fort, um Ihren Workload-Anforderungen gerecht zu werden.

Wenn Ihr Cluster gestartet wird, veröffentlicht der CloudWatch Agent die Systemmetriken, die Sie in der Konfigurationsdatei angegeben haben, auf CloudWatch.

## Anwendungsmetriken mit Amazon EMR 7.0.0 konfigurieren

Sie können den CloudWatch Amazon-Agenten so konfigurieren, dass er zusätzlich zu Systemmetriken auch Anwendungsmetriken für HDFS und YARN veröffentlicht. Gehen Sie wie folgt vor, um den Agent für die Veröffentlichung von Anwendungsmetriken zu konfigurieren:

1. Erstellen oder wählen Sie einen Bucket in Ihrem Amazon S3 S3-Konto aus, in dem Sie die Konfigurationsdateien speichern möchten, die die CloudWatch Agent-Metriken spezifizieren.
2. Navigieren Sie als Nächstes zum `aws-emr-utilities` Repo auf GitHub und laden Sie die folgenden Skripte herunter:
  - [`install\_app\_metrics\_launcher.sh`](#) – Ein Skript, das heruntergeladen und dann `install_app_metrics.sh` im Hintergrund ausführt, sodass der Knoten das Bootstrapping beenden kann.
  - [`install\_app\_metrics.sh`](#) – Ein Skript, das darauf wartet, dass die Instance, auf der es ausgeführt wird, das Bootstrapping abgeschlossen hat, und dann die Konfiguration in YAML-Dateien anwendet, die Sie in einem nächsten Schritt herunterladen werden.
3. Öffnen Sie jede Datei und ersetzen Sie `my-s3-bucket` durch den Namen Ihres Buckets aus Schritt 1.
4. Laden Sie als Nächstes die folgenden YAML-Mapping-Dateien herunter. Informationen darüber, wie diese YAML-Dateien strukturiert sind, finden Sie [javaagent](#) im OpenTelemetry Instrumentation for Java GitHub Repo.
  - [`datanode-metrics.yaml`](#) – Die Konfiguration für Hadoop DataNode-Metriken.
  - [`namenode-metrics.yaml`](#) – Die Konfiguration für Hadoop NameNode-Metriken.
  - [`nodemanager-metrics.yaml`](#) – Die Konfiguration für Yarn NodeManager-Metriken.
  - [`resourcemanager-metrics.yaml`](#) – Die Konfiguration für Yarn ResourceManager-Metriken.
5. Laden Sie die zwei SH- und vier YAML-Dateien in Ihren S3-Bucket hoch.
6. Jetzt können Sie zur Amazon EMR-Konsole navigieren und mit dem CloudWatch Agenten einen neuen Cluster erstellen. Wählen Sie im Navigationsbereich unter EMR in EC2 die Option Cluster und dann Create cluster (Cluster erstellen) aus.
7. Wählen Sie im Abschnitt Name and applications (Name und Anwendungen) Amazon-EMR-Version 7.0.0 oder höher aus.
8. Wählen Sie unter Anwendungspaket das Paket oder die benutzerdefinierte Gruppe von Apps aus, die Sie in Ihrem Cluster installieren möchten, und schließen Sie den CloudWatch Agenten in Ihre Auswahl ein.
9. Wählen Sie im Abschnitt Bootstrap actions (Bootstrap-Aktionen) die Option Add (Hinzufügen) aus.
  - Geben Sie für Name `install_app_metrics_launcher.sh` ein.



- Fügen Sie für Script location (Skript-Standort) **s3://my-s3-bucket/install\_app\_metrics\_launcher.sh** ein. Ersetzen Sie *my-s3-bucket* durch den Pfad zu Ihrem S3-Bucket.
- Lassen Sie den Argument-Block leer.

10. Wählen Sie Bootstrap-Aktion hinzufügen aus.

11. Fahren Sie mit der Erstellung des Clusters fort, um Ihren Workload-Anforderungen gerecht zu werden.

Wenn Ihr Cluster gestartet wird, veröffentlicht der CloudWatch Agent die Anwendungsmetriken, die Sie angegeben haben, zusammen mit den Systemmetriken für CloudWatch.

## Amazon Managed Service für Prometheus als Cloud-Speicher für Metriken mit Amazon EMR 7.0.0 konfigurieren

Sie können den CloudWatch Amazon-Agenten so konfigurieren, dass er Metriken stattdessen in Amazon Managed Service for Prometheus veröffentlicht. CloudWatch

### Note

Sie können Metriken vom CloudWatch Amazon-Agenten entweder auf Amazon Managed Service for Prometheus oder auf Amazon veröffentlichen CloudWatch, aber Sie können die Metriken nicht für beide Services für denselben Cluster veröffentlichen.

Um den Agenten für die Veröffentlichung von Metriken in Amazon Managed Service for Prometheus zu konfigurieren, müssen Sie die `aps:RemoteWrite` AWS Identity and Access Management (IAM) -Berechtigung zum Amazon EC2 EC2-Instance-Profil für Amazon EMR hinzufügen. Die folgende Beispielrichtlinie enthält die erforderliche Berechtigung:

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": "aps:RemoteWrite",
      "Resource": "*"
    }
  ]
}
```

```
}
```

Verwenden Sie den CloudWatch Agenten auf einem EMR-Cluster, um Metriken in Amazon Managed Service for Prometheus zu veröffentlichen

Sobald die Servicerichtlinie über die richtigen Berechtigungen verfügt, führen Sie die folgenden Schritte aus, um einen Cluster zu starten, der den CloudWatch Agenten verwendet, um Metriken in Amazon Managed Service for Prometheus zu veröffentlichen.

1. Verwenden Sie das AWS Management Console oder AWS CLI , um einen Amazon Managed Service for Prometheus Workspace zu erstellen. Weitere Informationen finden Sie unter [Einen Workspace erstellen](#) im Benutzerhandbuch zu Amazon Managed Service für Prometheus.
2. Erstellen Sie einen Bucket in Ihrem Amazon-S3-Konto oder wählen Sie einen aus, in dem Sie die Startdateien speichern möchten, die Amazon Managed Service für Prometheus als Cloud-Speicher angeben.
3. Navigieren Sie als Nächstes zum `aws-emr-utilities` Repo auf GitHub und laden Sie die folgenden Skripts herunter:
  - [add\\_prometheus\\_endpoint\\_launcher.sh](#) – Ein Skript, das heruntergeladen und dann `add_prometheus_endpoint.sh` im Hintergrund ausführt, sodass der Knoten das Bootstrapping beenden kann.
  - [add\\_prometheus\\_endpoint.sh](#)— Ein Skript, das darauf wartet, dass die Instance, auf der es ausgeführt wird, den Bootstrapping abgeschlossen hat, und dann den CloudWatch Agenten so konfiguriert, dass er auf dem Amazon Managed Service for Prometheus-Endpoint veröffentlicht, den Sie beim Starten Ihres Clusters als Argument angeben.
4. Öffnen Sie jede Datei und ersetzen Sie `my-s3-bucket` durch den Namen Ihres Buckets aus Schritt 2.
5. Verwenden Sie den AWS CLI , um einen EMR-Cluster mit der `add_prometheus_endpoint_launcher.sh` Bootstrap-Aktion zu erstellen. Ersetzen Sie `my-s3-bucket` im folgenden Befehl durch den Bucket, der die Bootstrap-Aktion enthält, und ersetzen Sie `managedpro-remote-write-workspace-url` durch den Remote-Write-Endpoint für Ihren Workspace von Amazon Managed Service für Prometheus. Stellen Sie sicher, dass Sie ein Amazon-EMR-Release-Label von `emr-7.0.0` oder höher angeben.

```
aws emr create-cluster --name managedpro-cluster \  
  --release-label emr-7.0.0 \  
  --applications Name=Hadoop Name=AmazonCloudWatchAgent \  
  --bootstrap-actions Name=add_prometheus_endpoint_launcher
```

```
--ec2-attributes KeyName=myKey --instance-type m7g.2xlarge \  
--instance-count 3 --use-default-roles  
--bootstrap-actions Name='Add Prometheus Endpoint',Path=s3://my-s3-bucket/  
add_prometheus_endpoint_launcher.sh,Args='managedpro-remote-write-workspace-url'
```

Wenn Ihr Cluster gestartet wird, veröffentlicht der CloudWatch Agent die gesammelten Metriken in Amazon Managed Service for Prometheus.

Amazon Managed Service für Prometheus als eine Datenquelle für Amazon Managed Grafana verwenden

Sobald Amazon EMR die Cluster-Metriken in Amazon Managed Service für Prometheus veröffentlicht hat, können Sie die Metriken mit Amazon Managed Grafana über die folgenden Schritten visualisieren:

1. Verwenden Sie den AWS Management Console , um einen Amazon Managed Grafana-Arbeitsbereich und einen Benutzer mit den entsprechenden Berechtigungen zu erstellen. Weitere Informationen finden Sie unter [Einen Workspace erstellen](#) im Benutzerhandbuch zu Amazon Managed Grafana.
2. Fügen Sie Ihren Workspace von Amazon Managed Service für Prometheus als eine Datenquelle zu Amazon Managed Grafana hinzu. Weitere Informationen finden Sie unter [AWS -Datenquellenkonfiguration verwenden, um Amazon Managed Service für Prometheus als Datenquelle hinzuzufügen](#) im Benutzerhandbuch zu Amazon Managed Grafana.

#### Note

Der CloudWatch Agent verfügt über einen Prometheus-Exporter, der bestimmte Attribute umbenennt. Für die Standardkennzahlen verwendet Amazon Managed Service for Prometheus Unterstriche anstelle der Punkte, die Amazon verwendet. CloudWatch Wenn Sie also Amazon Managed Grafana verwenden, um die Standardmetriken in Amazon Managed Service für Prometheus zu visualisieren, werden die Labels als `jobflow_id`, `instance_id` und `service_name` angezeigt.

Außerdem verwenden alle Anwendungsmetriken, die der CloudWatch Agent auf Amazon Managed Service for Prometheus veröffentlicht, das Label `job` anstelle von `service_name`. Systemmetriken verwenden jedoch weiterhin das Label `service_name`.

## Überlegungen und Einschränkungen

- Der native CloudWatch Amazon-Agent ist für Cluster verfügbar, die Sie mit Amazon EMR-Versionen 7.0.0 und höher erstellen.
- Die Konfiguration für den CloudWatch Agenten in Amazon EMR 7.0.0 erfordert die Verwendung von [Bootstrap-Aktionen](#). In einer kommenden Version wird Amazon EMR zusätzliche Konfigurationsoptionen über die Amazon-EMR-API bereitstellen.
- Sie können den Amazon CloudWatch EMR-Agenten nicht installieren, wenn Sie den CloudWatch Agenten bereits mit einer anderen Methode wie Agent (SSM AWS Systems Manager Agent) in der Region bereitgestellt haben, in der Sie Ihren Cluster erstellen. Dies führt dazu, dass Ihr Cluster mit Fehlern beendet wird.

## CloudWatch Versionsverlauf des Agenten

In der folgenden Tabelle sind die Versionen von aufgeführt, die in jeder Release-Version von Amazon EMR AmazonCloudWatchAgent enthalten sind, zusammen mit den Komponenten, die mit der Anwendung installiert wurden. Informationen zu den Komponentenversionen in den einzelnen Versionen finden Sie im Abschnitt Komponentenversion für Ihre Version in [Amazon-EMR-7.x-Versionen](#), [Amazon-EMR-6.x-Versionen](#) oder [Amazon-EMR-5.x-Versionen](#).

### AmazonCloudWatchAgent Versionsinformationen

Amazon-EMR-Version sbezeichnung	AmazonCloudWatchAgent Version	Komponenten, die installie rt wurden mit AmazonClo udWatchAgent
emr-7.1.0	1.300032,2	adot-java-agent, emrfs, emr- amazon-cloudwatch-agent , emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hd fs-datanode, hadoop-hdfs- library, hadoop-hdfs-nameno de, hadoop-httpfs-server, hadoop-kms-server, hadoop- mapred, hadoop-yarn-nodema nager, hadoop-yarn-resour

Amazon-EMR-Version sbezeichnung	AmazonCloudWatchAgent Version	Komponenten, die installie rt wurden mit AmazonClo udWatchAgent
		cemanager, hadoop-yarn- timeline-server
emr-7.0.0	1,300031,1	adot-java-agent, emr-amazon- cloudwatch-agent

# Delta Lake

Delta Lake ist ein Storage-Layer-Framework für Lakehouse-Architekturen, die üblicherweise auf Amazon S3 basieren. Mit Amazon-EMR-Versionen 6.9.0 und höher können Sie [Apache Spark 3.x](#) auf Amazon-EMR-Clustern mit Delta-Lake-Tabellen verwenden. Weitere Informationen zu Lakehouses mit Delta Lake finden Sie unter <https://delta.io/>.

In der folgenden Tabelle wird die Version von Amazon EMR, die in der neuesten Version von Amazon-EMR-7.x-Serie enthalten ist, zusammen mit den Komponenten aufgeführt, die Amazon EMR mit Delta installiert.

Informationen zur Version der Komponenten, die in dieser Version mit Delta installiert wurden, finden Sie unter [Komponentenversionen von Version 7.1.0](#).

Delta-Versionsinformationen für emr-7.1.0

Amazon-EMR-Version sbezeichnung	Delta-Version	Mit Delta installierte Komponenten
emr-7.1.0	Delta 3.0.0	Not available.

In der folgenden Tabelle wird die Version von Amazon EMR, die in der neuesten Version von Amazon-EMR-6.x-Serie enthalten ist, zusammen mit den Komponenten aufgeführt, die Amazon EMR mit Delta installiert.

Die Version der Komponenten, die mit Delta in dieser Version installiert wurden, finden Sie unter [Komponentenversionen der Version 6.15.0](#).

Delta-Versionsinformationen für emr-6.15.0

Amazon-EMR-Version sbezeichnung	Delta-Version	Mit Delta installierte Komponenten
emr-6.15.0	Delta 2.4.0	Not available.

# Einführung in Delta Lake

Delta Lake ist ein Open-Source-Projekt, das bei der Implementierung moderner Data-Lake-Architekturen hilft, die üblicherweise auf Amazon S3 basieren. Delta Lake bietet folgende Funktionen:

- Atomare, konsistente, isolierte, dauerhafte (ACID) Transaktionen auf Spark. Den Lesern wird während eines Spark-Auftrags eine konsistente Ansicht der Tabelle angezeigt.
- Skalierbares Handling von Metadaten mit verteilter Verarbeitung durch Spark.
- Kombiniert Streaming- und Batch-Anwendungsfälle mit derselben Delta-Tabelle.
- Automatische Durchsetzung von Schemas zur Vermeidung fehlerhafter Datensätze bei der Datenerfassung.
- Zeitreise mit der Versionsverwaltung von Daten.
- Unterstützt Zusammenführungs-, Aktualisierungs- und Löschvorgänge für komplexe Anwendungsfälle wie Change Data Capture (CDC), Streaming-Upserts und mehr.

## Einen Cluster verwenden, auf dem Delta Lake installiert ist

Themen

- [Verwenden Sie einen Delta Lake-Cluster mit Flink](#)
- [Einen Delta-Lake-Cluster mit Trino verwenden](#)
- [Verwenden Sie einen Delta-Lake-Cluster mit Spark](#)
- [Einen Delta-Lake-Cluster mit Spark und AWS Glue verwenden](#)

## Verwenden Sie einen Delta Lake-Cluster mit Flink

Mit Amazon-EMR-Version 6.11 und höher können Sie Delta Lake mit Ihrem Flink-Cluster verwenden. In den folgenden Beispielen wird der verwendet AWS CLI , um mit Delta Lake auf einem Amazon EMR Flink-Cluster zu arbeiten.

### Note

Amazon EMR unterstützt die DataStream Flink-API, wenn Sie Delta Lake mit einem Flink-Cluster verwenden.

## Einen Delta-Lake-Cluster erstellen

1. Erstellen Sie eine Datei, `delta_configurations.json`, mit folgendem Inhalt:

```
[{"Classification":"delta-defaults",  
  "Properties":{"delta.enabled":"true"}}]
```

2. Erstellen Sie einen Cluster mit der folgenden Konfiguration. Ersetzen Sie `example` Amazon S3 bucket path und subnet ID durch Ihre eigenen Werte.

```
aws emr create-cluster  
--release-label emr-6.11.0  
--applications Name=Flink  
--configurations file://delta_configurations.json  
--region us-east-1 --name My_Spark_Delta_Cluster  
--log-uri s3://DOC-EXAMPLE-BUCKET/  
--instance-type m5.xlarge  
--instance-count 3  
--service-role EMR_DefaultRole_V2  
--ec2-attributes  
  InstanceProfile=EMR_EC2_DefaultRole,SubnetId=subnet-1234567890abcdef0
```

## Eine Flink-Yarn-Sitzung installieren

Um eine Flink-Yarn-Sitzung zu initialisieren, führen Sie den folgenden Befehl aus:

```
flink-yarn-session -d
```

## Einen Flink-Auftrag mit Delta Lake erstellen

Die folgenden Beispiele zeigen, wie Sie mit SBT oder Maven für die Erstellung Ihres Flink-Auftrags mit Delta Lake verwenden.

sbt

[sbt](#) ist ein Build-Tool für Scala, das Sie mit wenig bis gar keiner Konfiguration verwenden können, wenn Sie kleine Projekte haben.

```
libraryDependencies += Seq(  
  "io.delta" %% "delta-flink" % deltaConnectorsVersion % "provided",
```



```
"io.delta" %% "delta-standalone" % deltaConnectorsVersion % "provided",
"org.apache.flink" %% "flink-clients" % flinkVersion % "provided",
"org.apache.flink" %% "flink-parquet" % flinkVersion % "provided",
"org.apache.hadoop" % "hadoop-client" % hadoopVersion % "provided",
"org.apache.flink" % "flink-table-common" % flinkVersion % "provided",
"org.apache.flink" %% "flink-table-runtime" % flinkVersion % "provided")
```

## Maven

[Maven](#) ist ein Open-Source-Tool zur Build-Automatisierung von der Apache Software Foundation. Mit Maven können Sie einen Flink-Auftrag mit Delta Lake auf Amazon EMR erstellen, veröffentlichen und bereitstellen.

```
<project>
<properties>
  <scala.main.version>2.12</scala.main.version>
  <delta-connectors-version>0.6.0</delta-connectors-version>
  <flink-version>1.16.1</flink-version>
  <hadoop-version>3.1.0</hadoop-version>
</properties>

<dependencies>
  <dependency>
    <groupId>io.delta</groupId>
    <artifactId>delta-flink</artifactId>
    <version>${delta-connectors-version}</version>
    <scope>provided</scope>
  </dependency>
  <dependency>
    <groupId>io.delta</groupId>
    <artifactId>delta-standalone_${scala-main-version}</artifactId>
    <version>${delta-connectors-version}</version>
    <scope>provided</scope>
  </dependency>
  <dependency>
    <groupId>org.apache.flink</groupId>
    <artifactId>flink-clients</artifactId>
    <version>${flink-version}</version>
    <scope>provided</scope>
  </dependency>
  <dependency>
    <groupId>org.apache.flink</groupId>
    <artifactId>flink-parquet</artifactId>
```

```

    <version>${flink-version}</version>
    <scope>provided</scope>
</dependency>
<dependency>
  <groupId>org.apache.hadoop</groupId>
  <artifactId>hadoop-client</artifactId>
  <version>${hadoop-version}</version>
  <scope>provided</scope>
</dependency>
<dependency>
  <groupId>org.apache.flink</groupId>
  <artifactId>flink-table-common</artifactId>
  <version>${flink-version}</version>
  <scope>provided</scope>
</dependency>
<dependency>
  <groupId>org.apache.flink</groupId>
  <artifactId>flink-table-runtime</artifactId>
  <version>${flink-version}</version>
  <scope>provided</scope>
</dependency>
</dependencies>

```

## Mit der Flink-DataStream-API in eine Delta-Tabelle schreiben

Verwenden Sie das folgende Beispiel, um eine DeltaSink zu erstellen, mit der Sie in die Tabelle schreiben können `deltaTablePath`:

```

public static DataStream<RowData> createDeltaSink(
    DataStream<RowData> stream,
    String deltaTablePath,
    RowType rowType) {
    Configuration configuration = new Configuration();
    DeltaSink<RowData> deltaSink = DeltaSink
        .forRowData(
            new org.apache.flink.core.fs.Path(deltaTablePath),
            configuration,
            rowType)
        .build();
    stream.sinkTo(deltaSink);
    return stream;
}

```

## Mit der Flink-DataStream-API aus einer Delta-Tabelle lesen

Verwenden Sie das folgende Beispiel, um eine Grenze zu erstellen `DeltaSource`, die aus der Tabelle mit einem gelesen werden soll `deltaTablePath`:

```
public static DataStream<RowData> createBoundedDeltaSourceAllColumns(
    StreamExecutionEnvironment env,
    String deltaTablePath) {
    Configuration configuration = new Configuration();
    DeltaSource<RowData> deltaSource = DeltaSource
        .forBoundedRowData(
            new org.apache.flink.core.fs.Path(deltaTablePath),
            configuration)
        .build();

    return env.fromSource(deltaSource, WatermarkStrategy.noWatermarks(), "delta-
source");
}
```

## Sink-Erstellung mit Multi-Cluster-Unterstützung für Delta Lake Standalone

Verwenden Sie das folgende Beispiel, um eine Tabelle `DeltaSink` zum Schreiben in eine Tabelle mit `A deltaTablePath` - und [Multi-Cluster-Unterstützung](#) zu erstellen:

```
public DataStream<RowData> createDeltaSink(
    DataStream<RowData> stream,
    String deltaTablePath) {
    Configuration configuration = new Configuration();
    configuration.set("spark.delta.logStore.s3.impl",
"io.delta.storage.S3DynamoDBLogStore");
    configuration.set("spark.io.delta.storage.S3DynamoDBLogStore.ddb.tableName",
"delta_log");
    configuration.set("spark.io.delta.storage.S3DynamoDBLogStore.ddb.region", "us-
east-1");

    DeltaSink<RowData> deltaSink = DeltaSink
        .forRowData(
            new Path(deltaTablePath),
            configuration,
            rowType)
        .build();
    stream.sinkTo(deltaSink);
}
```

```
    return stream;
}
```

## Den Flink-Auftrag ausführen

Den folgenden Befehl verwenden, um Ihren Job auszuführen:

```
flink run FlinkJob.jar
```

## Einen Delta-Lake-Cluster mit Trino verwenden

Mit Amazon-EMR-Versionen 6.9.0 und höher können Sie Delta Lake mit Ihrem Trino-Cluster verwenden.

In diesem Tutorial werden wir den Trino-Cluster verwenden AWS CLI , um mit Delta Lake auf Amazon EMR zu arbeiten.

### Einen Delta-Lake-Cluster erstellen

1. Erstellen Sie eine Datei, `delta_configurations.json`, und legen Sie Werte für den ausgewählten Katalog fest. Wenn Sie beispielsweise den Hive-Metastore als Ihren Katalog verwenden möchten, sollte Ihre Datei den folgenden Inhalt haben:

```
[{"Classification":"delta-defaults",
  "Properties":{"delta.enabled":"true"}},
 {"Classification":"trino-connector-delta",
  "Properties":{"hive.metastore.uri":"thrift://localhost:9083"}}]
```

Wenn Sie den AWS Glue-Katalog als Ihren Shop verwenden möchten, sollte Ihre Datei den folgenden Inhalt haben:

```
[{"Classification":"delta-defaults",
  "Properties":{"delta.enabled":"true"}},
 {"Classification":"trino-connector-delta",
  "Properties":{"hive.metastore":"glue"}}]
```

2. Erstellen Sie einen Cluster mit der folgenden Konfiguration und ersetzen Sie **example Amazon S3 bucket path** und **subnet ID** durch Ihre eigenen.

```
aws emr create-cluster
```

```
--release-label emr-6.9.0
--applications Name=Trino
--configurations file://delta_configurations.json
--region us-east-1 --name My_Spark_Delta_Cluster
--log-uri s3://DOC-EXAMPLE-BUCKET/
--instance-type m5.xlarge
--instance-count 2
--service-role EMR_DefaultRole_V2
--ec2-attributes
InstanceProfile=EMR_EC2_DefaultRole,SubnetId=subnet-1234567890abcdef0
```

## Initialisieren Sie die Trino-Sitzung für Delta Lake

Um die Trino-Sitzung zu initialisieren, führen Sie den folgenden Befehl aus

```
trino-cli --catalog delta
```

## In eine Delta-Lake-Tabelle schreiben

Erstellen Sie Ihre Tabelle mit den folgenden SQL-Befehlen und schreiben Sie in sie:

```
SHOW SCHEMAS;

CREATE TABLE default.delta_table (id int, data varchar, category varchar) WITH
( location = 's3://DOC-EXAMPLE-BUCKET/<prefix>');

INSERT INTO default.delta_table VALUES (1, 'a', 'c1'), (2, 'b', 'c2'), (3, 'c', 'c3');
```

## Von einer Delta-Lake-Tabelle lesen

Lesen Sie mit dem folgenden SQL-Befehl aus Ihrer Tabelle:

```
SELECT * from default.delta_table;
```

## Verwenden Sie einen Delta-Lake-Cluster mit Spark

Ab Amazon-EMR-Version 6.9.0 können Sie Delta Lake mit Ihrem Spark-Cluster verwenden, ohne dass Bootstrap-Aktionen erforderlich sind. Für Amazon-EMR-Versionen 6.8.0 und niedriger können Sie Bootstrap-Aktionen verwenden, um die erforderlichen Abhängigkeiten vorab zu installieren.

In den folgenden Beispielen wird der verwendet AWS CLI , um mit Delta Lake auf einem Amazon EMR Spark-Cluster zu arbeiten.

Um Delta Lake auf Amazon EMR mit dem zu verwenden AWS Command Line Interface, erstellen Sie zunächst einen Cluster. Informationen zur Angabe der Delta Lake-Klassifizierung mit finden Sie unter [Bereitstellen einer Konfiguration mithilfe von AWS Command Line Interface, AWS Command Line Interface wenn Sie einen Cluster erstellen oder Bereitstellen einer Konfiguration mit dem Java SDK, wenn Sie einen Cluster erstellen](#).

1. Erstellen Sie eine Datei, `configurations.json`, mit folgendem Inhalt:

```
[{"Classification":"delta-defaults", "Properties":{"delta.enabled":"true"} ]]
```

2. Erstellen Sie einen Cluster mit der folgenden Konfiguration und ersetzen Sie die Beispiele **bucket path** und **subnet ID** in Amazon S3 durch Ihre eigenen.

```
aws emr create-cluster
  --release-label emr-6.9.0
  --applications Name=Spark
  --configurations file://delta_configurations.json
  --region us-east-1
  --name My_Spark_Delta_Cluster
  --log-uri s3://DOC-EXAMPLE-BUCKET/
  --instance-type m5.xlarge
  --instance-count 2
  --service-role EMR_DefaultRole_V2
  --ec2-attributes
  InstanceProfile=EMR_EC2_DefaultRole,SubnetId=subnet-1234567890abcdef0
```

Alternativ können Sie einen Amazon-EMR-Cluster und eine Spark-Anwendung mit den folgenden Dateien als JAR-Abhängigkeiten in einem Spark-Job erstellen:

```
/usr/share/aws/delta/lib/delta-core.jar,
/usr/share/aws/delta/lib/delta-storage.jar,
```

```
/usr/share/aws/delta/lib/delta-storage-s3-dynamodb.jar
```

### Note

Wenn Sie Amazon EMR-Versionen 6.9.0 oder höher verwenden, verwenden Sie `/usr/share/aws/delta/lib/delta-spark.jar` anstelle von `/usr/share/aws/delta/lib/delta-core.jar`

Weitere Informationen finden Sie unter [Anwendungen einreichen](#).

Um eine JAR-Abhängigkeit in den Spark-Auftrag aufzunehmen, können Sie der Spark-Anwendung die folgenden Konfigurationseigenschaften hinzufügen:

```
--conf "spark.jars=/usr/share/aws/delta/lib/delta-core.jar,  
/usr/share/aws/delta/lib/delta-storage.jar,  
/usr/share/aws/delta/lib/delta-storage-s3-dynamodb.jar"
```

Weitere Informationen zu den Abhängigkeiten von Spark-Aufträgen erhalten Sie unter [Abhängigkeitsmanagement](#).

Wenn Sie Amazon EMR-Versionen 6.9.0 oder höher verwenden, fügen Sie stattdessen die `/usr/share/aws/delta/lib/delta-spark.jar` Konfiguration hinzu.

```
--conf "spark.jars=/usr/share/aws/delta/lib/delta-spark.jar,  
/usr/share/aws/delta/lib/delta-storage.jar,  
/usr/share/aws/delta/lib/delta-storage-s3-dynamodb.jar"
```

## Eine Spark-Sitzung für Delta Lake initialisieren

Die folgenden Beispiele zeigen, wie Sie die interaktive Spark-Shell starten, Spark Submit verwenden oder Amazon EMR Notebooks verwenden, um mit Delta Lake auf Amazon EMR zu arbeiten.

spark-shell

1. Verbinden Sie sich dem Primärknoten über SSH. Weitere Informationen finden Sie unter [Stellen Sie über SSH eine Verbindung zum Primärknoten her](#) im Verwaltungshandbuch für Amazon EMR.

2. Geben Sie den folgenden Befehl ein, um die Spark-Shell zu starten. Um die PySpark Shell zu verwenden, ersetzen Sie `durchspark-shell` durch `pyspark`

```
spark-shell \  
  --conf "spark.sql.extensions=io.delta.sql.DeltaSparkSessionExtension" \  
  --conf  
  "spark.sql.catalog.spark_catalog=org.apache.spark.sql.delta.catalog.DeltaCatalog"
```

## spark-submit

1. Verbinden Sie sich dem Primärknoten über SSH. Weitere Informationen finden Sie unter [Stellen Sie über SSH eine Verbindung zum Primärknoten her](#) im Verwaltungshandbuch für Amazon EMR.
2. Geben Sie den folgenden Befehl ein, um die Spark-Sitzung für Delta Lake zu starten.

```
spark-submit  
  -conf "spark.sql.extensions=io.delta.sql.DeltaSparkSessionExtension"  
  -conf  
  "spark.sql.catalog.spark_catalog=org.apache.spark.sql.delta.catalog.DeltaCatalog"
```

## EMR Studio notebooks

Um eine Spark-Sitzung mit Amazon-EMR-Studio-Notebooks zu initialisieren, konfigurieren Sie Ihre Spark-Sitzung mithilfe des `%%configure` magischen Befehls in Ihrem Amazon EMR Notebook, wie im folgenden Beispiel. Weitere Informationen finden Sie unter [Verwendung von EMR-Notebooks-Magics](#) im Verwaltungshandbuch für Amazon EMR.

```
%%configure -f  
{  
  "conf": {  
    "spark.sql.extensions": "io.delta.sql.DeltaSparkSessionExtension",  
    "spark.sql.catalog.spark_catalog":  
    "org.apache.spark.sql.delta.catalog.DeltaCatalog"  
  }  
}
```



## In eine Delta-Lake-Tabelle schreiben

Das folgende Beispiel zeigt, wie ein Delta Lake-Dataset erstellt DataFrame und als Delta Lake-Datensatz geschrieben wird. Das Beispiel zeigt, wie Sie mit Datensätzen mit der Spark-Shell arbeiten, während Sie mit dem Primärknoten verbunden sind und SSH als Standard-Hadoop-Benutzer verwenden.

### Note

Um Codebeispiele in die Spark-Shell einzufügen, geben Sie an der Eingabeaufforderung `:paste` ein, fügen das Beispiel ein und drücken dann CTRL + D.

## PySpark

Spark umfasst auch eine auf Python basierende Shell, `pyspark`, mit der Sie Prototypen von in Python geschriebenen Spark-Programmen entwickeln können. Rufen Sie genau wie bei `spark-shell`, `pyspark` auf dem Primärknoten auf.

```
## Create a DataFrame
data = spark.createDataFrame([("100", "2015-01-01", "2015-01-01T13:51:39.340396Z"),
                              ("101", "2015-01-01", "2015-01-01T12:14:58.597216Z"),
                              ("102", "2015-01-01", "2015-01-01T13:51:40.417052Z"),
                              ("103", "2015-01-01", "2015-01-01T13:51:40.519832Z")],
                              ["id", "creation_date", "last_update_time"])

## Write a DataFrame as a Delta Lake dataset to the S3 location
spark.sql("""CREATE TABLE IF NOT EXISTS delta_table (id string, creation_date
string,
last_update_time string)
USING delta location
's3://DOC-EXAMPLE-BUCKET/example-prefix/db/delta_table'""");

data.writeTo("delta_table").append()
```

## Scala

```
import org.apache.spark.sql.SaveMode
import org.apache.spark.sql.functions._

// Create a DataFrame
```

```
val data = Seq(("100", "2015-01-01", "2015-01-01T13:51:39.340396Z"),
("101", "2015-01-01", "2015-01-01T12:14:58.597216Z"),
("102", "2015-01-01", "2015-01-01T13:51:40.417052Z"),
("103", "2015-01-01", "2015-01-01T13:51:40.519832Z")).toDF("id", "creation_date",
"last_update_time")

// Write a DataFrame as a Delta Lake dataset to the S3 location
spark.sql("""CREATE TABLE IF NOT EXISTS delta_table (id string,
creation_date string,
last_update_time string)
USING delta location
's3://DOC-EXAMPLE-BUCKET/example-prefix/db/delta_table'""");

data.write.format("delta").mode("append").saveAsTable("delta_table")
```

## SQL

```
-- Create a Delta Lake table with the S3 location
CREATE TABLE delta_table(id string,
creation_date string,
last_update_time string)
USING delta LOCATION
's3://DOC-EXAMPLE-BUCKET/example-prefix/db/delta_table';

-- insert data into the table
INSERT INTO delta_table VALUES ("100", "2015-01-01",
"2015-01-01T13:51:39.340396Z"),
("101", "2015-01-01", "2015-01-01T12:14:58.597216Z"),
("102", "2015-01-01", "2015-01-01T13:51:40.417052Z"),
("103", "2015-01-01", "2015-01-01T13:51:40.519832Z");
```

## Von einer Delta-Lake-Tabelle lesen

### PySpark

```
ddf = spark.table("delta_table")
ddf.show()
```

### Scala

```
val ddf = spark.table("delta_table")
```

```
ddf.show()
```

## SQL

```
SELECT * FROM delta_table;
```

## Einen Delta-Lake-Cluster mit Spark und AWS Glue verwenden

Um den AWS Glue-Katalog als Metastore für Delta Lake-Tabellen zu verwenden, erstellen Sie einen Cluster mit den folgenden Schritten. Informationen zur Angabe der Delta Lake-Klassifizierung mithilfe von finden Sie unter [Bereitstellen einer Konfiguration mithilfe von AWS Command Line Interface](#), [AWS Command Line Interface wenn Sie einen Cluster erstellen](#) oder [Bereitstellen einer Konfiguration mithilfe des Java SDK, wenn Sie einen Cluster erstellen](#).

### Einen Delta-Lake-Cluster erstellen

1. Erstellen Sie eine Datei, `configurations.json`, mit folgendem Inhalt:

```
[{"Classification":"delta-defaults",
  "Properties":{"delta.enabled":"true"}},
{"Classification":"spark-hive-site",
  "Properties":
  {"hive.metastore.client.factory.class":"com.amazonaws.glue.catalog.metastore.AWSGlueDataCat
```

2. Erstellen Sie einen Cluster mit der folgenden Konfiguration und ersetzen Sie **example Amazon S3 bucket path** und **subnet ID** durch Ihre eigenen.

```
aws emr create-cluster
  --release-label emr-6.9.0
  --applications Name=Spark
  --configurations file://delta_configurations.json
  --region us-east-1
  --name My_Spark_Delta_Cluster
  --log-uri s3://DOC-EXAMPLE-BUCKET/
  --instance-type m5.xlarge
  --instance-count 2
  --service-role EMR_DefaultRole_V2
```

```
--ec2-attributes
InstanceProfile=EMR_EC2_DefaultRole,SubnetId=subnet-1234567890abcdef0
```

## Überlegungen und Einschränkungen

- Delta Lake wird für die Verwendung mit Amazon-EMR-Versionen 6.9.0 und höher unterstützt. Sie können [Apache Spark 3.x](#) auf Amazon-EMR-Clustern mit Delta-Tabellen verwenden.
- Wir empfehlen, das S3-URI-Schema für S3-Standortpfade anstelle von s3a zu verwenden, um optimale Leistung, Sicherheit und Zuverlässigkeit zu erzielen. Weitere Informationen finden Sie unter [Arbeiten mit Speicher- und Dateisystemen](#).
- Mit Amazon EMR 7.0 werden Delta Universal Format (UniForm) und Convert-to-Iceberg-Anweisungen nicht unterstützt.
- Wenn Sie mit Amazon EMR 6.9 und 6.10 Delta-Lake-Tabellendaten in Amazon S3 speichern, werden die Spaltendaten NULL nach dem Umbenennungsvorgang der Spalten. Ab Amazon EMR 6.11 ist dieses Problem behoben. Weitere Informationen zum experimentellen Vorgang zum Umbenennen von Spalten finden Sie unter [Vorgang zum Umbenennen von Spalten](#) im Delta-Lake-Benutzerhandbuch.
- Wenn Sie außerhalb von Apache Spark eine Datenbank im AWS Glue-Datenkatalog erstellen, könnte die Datenbank ein leeres LOCATION Feld enthalten. Da Spark das Erstellen von Datenbanken mit einer leeren Standorteigenschaft nicht zulässt, erhalten Sie die folgende Fehlermeldung, wenn Sie Spark in Amazon EMR verwenden, um eine Delta-Tabelle in einer Glue-Datenbank zu erstellen und die Datenbank eine leere LOCATION-Eigenschaft hat:

```
IllegalArgumentException: Can not create a Path from an empty string
```

Um dieses Problem zu beheben, erstellen Sie die Datenbank im Datenkatalog mit einem gültigen, nicht leeren Pfad für das LOCATION-Feld. Schritte zur Implementierung dieser Lösung finden Sie unter [Unzulässige Argumentausnahme beim Erstellen einer Tabelle](#) im Amazon-Athena-Benutzerhandbuch.

## Versionsverlauf von Delta

In der folgenden Tabelle sind die Version von Delta aufgeführt, die in jeder Release-Version von Amazon EMR enthalten ist, zusammen mit den Komponenten, die mit der Anwendung installiert wurden. Informationen zu den Komponentenversionen in den einzelnen Versionen finden Sie im

Abschnitt Komponentenversion für Ihre Version in [Amazon-EMR-7.x-Versionen](#), [Amazon-EMR-6.x-Versionen](#) oder [Amazon-EMR-5.x-Versionen](#).

### Die Delta Versionsinformationen

Amazon-EMR-Version sbezeichnung	Delta-Version	Mit Delta installierte Komponenten
emr-7.1.0	3.0.0	Not available.
emr-7.0.0	3.0.0	Not available.
emr-6.15.0	2.4.0	Not available.
emr-6.14.0	2.4.0	Not available.
emr-6.13.0	2.4.0	Not available.
emr-6.12.0	2.4.0	Not available.
emr-6.11.1	2.2.0	Not available.
emr-6.11.0	2.2.0	Not available.
emr-6.10.1	2.2.0	Not available.
emr-6.10.0	2.2.0	Not available.
emr-6.9.1	2.1.0	Not available.
emr-6.9.0	2.1.0	Not available.

# Apache Flink

[Apache Flink](#) ist eine Streaming-Datenfluss-Engine, die mittels Datenquellen mit hohem Durchsatz die Echtzeitverarbeitung von Streams ermöglicht. Flink unterstützt Ereigniszeitsemantik für out-of-order Ereignisse, Exactly-Once-Semantik, Backpressure Control und APIs, die für das Schreiben von Streaming- und Batch-Anwendungen optimiert sind.

Außerdem verfügt Flink über Konnektoren für Datenquellen von Drittanbieter, wie z. B. die folgenden:

- [Amazon Kinesis Data Streams](#)
- [Apache Kafka](#)
- [Flink-Elasticsearch-Konnektor](#)
- [Twitter-Streaming-API](#)
- [Cassandra](#)

Amazon EMR unterstützt Flink als YARN-Anwendung, sodass Sie Ressourcen zusammen mit anderen Anwendungen in einem Cluster verwalten können. Flink-on-YARN ermöglicht das Senden vorübergehender Flink-Aufträge. Sie können auch einen langlebigen Cluster erstellen, der mehrere Aufträge akzeptiert und Ressourcen entsprechend der allgemeinen YARN-Reservierung zuweist.

Flink ist in Amazon EMR Version 5.1.0 und höher enthalten.

## Note

Support für die `FlinkKinesisConsumer`-Klasse wurde mit Amazon-EMR-Version 5.2.1 bereitgestellt.

Die folgende Tabelle listet die Version von Flink auf, die in der neuesten Version der Amazon-EMR-7.x-Serie enthalten ist, zusammen mit den Komponenten, die Amazon EMR mit Flink installiert.

[Die Version der Komponenten, die in dieser Version mit Flink installiert wurden, finden Sie unter Komponentenversionen der Version 7.1.0.](#)

## Flink-Versionsinformationen für emr-7.1.0

Amazon-EMR-Version sbezeichnung	Flink-Version	Mit Flink installierte Komponenten
emr-7.1.0	Flink 1.18.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config, hudi, delta

Die folgende Tabelle listet die Version von Flink auf, die in der neuesten Version der Amazon-EMR-6.x-Serie enthalten ist, zusammen mit den Komponenten, die Amazon EMR mit Flink installiert.

Die Version der Komponenten, die mit Flink in dieser Version installiert wurden, finden Sie unter [Komponentenversionen der Version 6.15.0](#).

## Flink-Versionsinformationen für emr-6.15.0

Amazon-EMR-Version sbezeichnung	Flink-Version	Mit Flink installierte Komponenten
emr-6.15.0	Flink 1.17.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client,

Amazon-EMR-Version sbezeichnung	Flink-Version	Mit Flink installierte Komponenten
		flink-jobmanager-config, hudi, delta-standalone-connectors

Die folgende Tabelle listet die Version von Flink auf, die in der neuesten Version der Amazon-EMR-5.x-Serie enthalten ist, zusammen mit den Komponenten, die Amazon EMR mit Flink installiert.

[Informationen zur Version der Komponenten, die in dieser Version mit Flink installiert wurden, finden Sie unter Komponentenversionen von Version 5.36.2.](#)

Flink-Versionsinformationen für emr-5.36.2

Amazon-EMR-Version sbezeichnung	Flink-Version	Mit Flink installierte Komponenten
emr-5.36.2	Flink 1.14.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config

## Themen

- [Erstellen eines Clusters mit Flink](#)
- [Konfiguration von Flink in Amazon EMR](#)
- [Arbeiten mit Flink-Jobs in Amazon EMR](#)
- [Verwenden der Scala-Shell](#)
- [Suchen der Flink Webschnittstelle](#)
- [Flink Autoscaler](#)



- [Optimieren der Neustartzeiten von Aufträgen für die Aufgabenwiederherstellung und -skalierung](#)
- [Arbeiten mit Flink-Aufträgen von Zeppelin in Amazon EMR](#)
- [Versionsverlauf von Flink](#)

## Erstellen eines Clusters mit Flink

Sie können einen Cluster mit dem AWS Management Console, AWS CLI, oder einem AWS SDK starten.

So starten Sie einen Cluster mit Flink über die Konsole

1. Öffnen Sie die Amazon EMR-Konsole unter <https://console.aws.amazon.com/emr>.
2. Wählen Sie Create Cluster (Cluster erstellen) und Go to advanced options (Zu erweiterten Optionen) aus.
3. Wählen Sie für das Feld Software Configuration (Softwarekonfiguration) die Option EMR Release emr-5.1.0 oder höher aus.
4. Wählen Sie Flink als Anwendung zusammen mit anderen zu installierenden Anwendungen aus.
5. Wählen Sie nach Bedarf weitere Optionen und anschließend Create cluster (Cluster erstellen) aus.

Um einen Cluster mit Flink von der AWS CLI

- Erstellen Sie den Cluster mit dem folgenden Befehl:

```
aws emr create-cluster --release-label emr-7.1.0 \  
--applications Name=Flink \  
--configurations file:///./configurations.json \  
--region us-east-1 \  
--log-uri s3://myLogUri \  
--instance-type m5.xlarge \  
--instance-count 2 \  
--service-role EMR_DefaultRole_V2 \  
--ec2-attributes KeyName=MyKeyName,InstanceProfile=EMR_EC2_DefaultRole \  
--steps Type=CUSTOM_JAR,Jar=command-runner.jar,Name=Flink_Long_Running_Session,\  
Args=flink-yarn-session,-d
```

**Note**

Linux-Zeilenumbruchzeichen (\) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (^).

## Konfiguration von Flink in Amazon EMR

### Konfigurieren Sie Flink mit Hive Metastore und Glue Catalog

Amazon EMR-Versionen 6.9.0 und höher unterstützen sowohl Hive Metastore als auch AWS Glue Catalog mit dem Apache Flink-Connector zu Hive. In diesem Abschnitt werden die Schritte beschrieben, die zur Konfiguration von [AWS Glue Catalog](#) und [Hive Metastore](#) mit Flink erforderlich sind.

#### Themen

- [Verwenden Sie den Hive Metastore](#)
- [Verwenden Sie den AWS Glue-Datenkatalog](#)

### Verwenden Sie den Hive Metastore

1. Erstellen Sie einen EMR-Cluster mit Version 6.9.0 oder höher und mindestens zwei Anwendungen: Hive und Flink.
2. Verwenden Sie [Script Runner](#), um das folgende Skript als Schrittfunktion auszuführen:

```
hive-metastore-setup.sh
```

```
sudo cp /usr/lib/hive/lib/antlr-runtime-3.5.2.jar /usr/lib/flink/lib
sudo cp /usr/lib/hive/lib/hive-exec-3.1.3*.jar /lib/flink/lib
sudo cp /usr/lib/hive/lib/libfb303-0.9.3.jar /lib/flink/lib
sudo cp /usr/lib/flink/opt/flink-connector-hive_2.12-1.15.2.jar /lib/flink/lib
sudo chmod 755 /usr/lib/flink/lib/antlr-runtime-3.5.2.jar
sudo chmod 755 /usr/lib/flink/lib/hive-exec-3.1.3*.jar
sudo chmod 755 /usr/lib/flink/lib/libfb303-0.9.3.jar
sudo chmod 755 /usr/lib/flink/lib/flink-connector-hive_2.12-1.15.2.jar
```

**Add step** ✕

**Step type** Custom JAR

**Name\***

**JAR location\***  JAR location maybe a path into S3 or a fully qualified java class in the classpath.

**Arguments**  These are passed to the main function in the JAR. If the JAR does not specify a main class in its manifest file you can specify another class name as the first argument.

**Action on failure**  What happens if the step fails

Cancel
Save

## Verwenden Sie den AWS Glue-Datenkatalog

1. Erstellen Sie einen EMR-Cluster mit Version 6.9.0 oder höher und mindestens zwei Anwendungen: Hive und Flink.
2. Wählen Sie in den Einstellungen des AWS Glue Data Catalogs die Option Für Hive-Tabellenmetadaten verwenden aus, um den Data Catalog im Cluster zu aktivieren.
3. Verwenden Sie [Script-Runner](#), um das folgende Skript als Schrittfunktion auszuführen: [Befehle und Skripts auf einem Amazon-EMR-Cluster ausführen](#):

glue-catalog-setup.sh

```

sudo cp /usr/lib/hive/auxlib/aws-glue-datacatalog-hive3-client.jar /usr/lib/flink/lib
sudo cp /usr/lib/hive/lib/antlr-runtime-3.5.2.jar /usr/lib/flink/lib
sudo cp /usr/lib/hive/lib/hive-exec-3.1.3*.jar /lib/flink/lib
sudo cp /usr/lib/hive/lib/libfb303-0.9.3.jar /lib/flink/lib
sudo cp /usr/lib/flink/opt/flink-connector-hive_2.12-1.15.2.jar /lib/flink/lib
sudo chmod 755 /usr/lib/flink/lib/aws-glue-datacatalog-hive3-client.jar
sudo chmod 755 /usr/lib/flink/lib/antlr-runtime-3.5.2.jar
sudo chmod 755 /usr/lib/flink/lib/hive-exec-3.1.3*.jar
sudo chmod 755 /usr/lib/flink/lib/libfb303-0.9.3.jar
sudo chmod 755 /usr/lib/flink/lib/flink-connector-hive_2.12-1.15.2.jar

```

**Add step**
✕

**Step type** Custom JAR

**Name\***

**JAR location\***  JAR location maybe a path into S3 or a fully qualified java class in the classpath.

**Arguments**  These are passed to the main function in the JAR. If the JAR does not specify a main class in its manifest file you can specify another class name as the first argument.

**Action on failure**  What happens if the step fails

Cancel
Save

## Konfigurieren Sie Flink mit einer Konfigurationsdatei

Sie können die Amazon-EMR-Konfigurations-API verwenden, um Flink mit einer Konfigurationsdatei zu konfigurieren. Folgende Dateien sind in der API konfigurierbar:

- `flink-conf.yaml`
- `log4j.properties`
- `flink-log4j-session`
- `log4j-cli.properties`

Die Hauptkonfigurationsdatei für Flink heißt `flink-conf.yaml`.

So konfigurieren Sie die Anzahl der Aufgaben-Slots für Flink mithilfe der AWS CLI

1. Erstellen Sie eine Datei, `configurations.json`, mit folgendem Inhalt:

```
[
  {
    "Classification": "flink-conf",
    "Properties": {
      "taskmanager.numberOfTaskSlots": "2"
    }
  }
]
```

]

2. Im nächsten Schritt erstellen Sie einen Cluster mit der folgenden Konfiguration:

```
aws emr create-cluster --release-label emr-7.1.0 \  
--applications Name=Flink \  
--configurations file://./configurations.json \  
--region us-east-1 \  
--log-uri s3://myLogUri \  
--instance-type m5.xlarge \  
--instance-count 2 \  
--service-role EMR_DefaultRole_V2 \  
--ec2-attributes KeyName=YourKeyName,InstanceProfile=EMR_EC2_DefaultRole
```

### Note

Sie können einige Konfigurationen auch mit der Flink-API ändern. Weitere Informationen finden Sie unter [Konzepte](#) in der Flink-Dokumentation.

Ab Amazon-EMR-Version 5.21.0 können Sie Cluster-Konfigurationen überschreiben und zusätzliche Konfigurationsklassifikationen für jede Instance-Gruppe in einem ausgeführten Cluster angeben. Dazu verwenden Sie die Amazon EMR-Konsole, das AWS Command Line Interface (AWS CLI) oder das AWS SDK. Weitere Informationen finden Sie unter [Angeben einer Konfiguration für eine Instance-Gruppe in einem aktiven Cluster](#).

## Parallelitätsoptionen

Als Eigentümer Ihrer Anwendung wissen Sie am besten, welche Ressourcen Aufgaben innerhalb von Flink zugewiesen werden müssen. Für Beispiele in dieser Dokumentation verwenden Sie die gleiche Anzahl von Aufgaben wie die Aufgaben-Instances, die Sie für die Anwendung nutzen. Wir empfehlen diese Vorgehensweise generell für die anfängliche Parallelität. Sie können jedoch die Granularität der Parallelität mithilfe von Aufgaben-Slots erhöhen. Dabei sollte die Anzahl von [virtuellen Cores](#) pro Instance im Allgemeinen nicht überschritten werden. Weitere Informationen über die Architektur von Flink finden Sie unter [Konzepte](#) in der Flink-Dokumentation.

## Konfigurieren von Flink auf einem EMR-Cluster mit mehreren Primärknoten

Der JobManager von Flink bleibt während des Failover-Prozesses für den primären Knoten in einem Amazon EMR-Cluster mit mehreren Primärknoten verfügbar. Ab Amazon EMR 5.28.0 wird

JobManager Hochverfügbarkeit auch automatisch aktiviert. Es ist keine manuelle Konfiguration erforderlich.

Bei Amazon EMR-Versionen 5.27.0 oder früher JobManager handelt es sich um einen einzigen Fehlerpunkt. Wenn der JobManager fehlschlägt, gehen alle Job-Status verloren und die laufenden Jobs werden nicht wieder aufgenommen. Sie können JobManager Hochverfügbarkeit aktivieren, indem Sie die Anzahl der Anwendungsversuche, das Checkpointing und die Aktivierung ZooKeeper als Statusspeicher für Flink konfigurieren, wie das folgende Beispiel zeigt:

```
[
  {
    "Classification": "yarn-site",
    "Properties": {
      "yarn.resourcemanager.am.max-attempts": "10"
    }
  },
  {
    "Classification": "flink-conf",
    "Properties": {
      "yarn.application-attempts": "10",
      "high-availability": "zookeeper",
      "high-availability.zookeeper.quorum": "%{hiera('hadoop:zk')}",
      "high-availability.storageDir": "hdfs:///user/flink/recovery",
      "high-availability.zookeeper.path.root": "/flink"
    }
  }
]
```

Sie müssen sowohl die maximalen Anwendungsmasterversuche für YARN als auch die maximalen Anwendungsversuche für Flink konfigurieren. Weitere Informationen finden Sie unter [Konfiguration der Hochverfügbarkeit des YARN-Clusters](#). Möglicherweise möchten Sie auch das Flink-Checkpointing so konfigurieren, dass neu gestartete Jobs von zuvor abgeschlossenen Checkpoints JobManager wiederhergestellt werden. Weitere Informationen finden Sie unter [Flink-Checkpointing](#).

## Konfiguration der Größe des Speicherprozesses

Für Amazon EMR-Versionen, die Flink 1.11.x verwenden, müssen Sie die Gesamtspeicherprozessgröße sowohl für () als auch für JobManager (`jobmanager.memory.process.size`) in konfigurieren. TaskManager `taskmanager.memory.process.size` `flink-conf.yaml` Sie können diese Werte festlegen,

indem Sie entweder den Cluster mit der Konfigurations-API konfigurieren oder diese Felder manuell über SSH auskommentieren. Flink bietet die folgenden Standardwerte.

- `jobmanager.memory.process.size`: 1 600 m
- `taskmanager.memory.process.size`: 1 728 m

Um JVM-Metaspace und Overhead auszuschließen, verwenden Sie die gesamte Flink-Speichergröße (`taskmanager.memory.flink.size`) anstelle von `taskmanager.memory.process.size`. Der Standardwert von `taskmanager.memory.process.size` beträgt 1 280 m. Es wird nicht empfohlen, sowohl `taskmanager.memory.process.size` als auch `taskmanager.memory.process.size` zu setzen.

Alle Amazon-EMR-Versionen, die Flink 1.12.0 und höher verwenden, haben die im Open-Source-Set für Flink aufgeführten Standardwerte als Standardwerte für Amazon EMR, sodass Sie sie nicht selbst konfigurieren müssen.

## Größe der Protokollausgabedatei konfigurieren

Flink-Anwendungscontainer erstellen drei Arten von Protokolldateien und schreiben in sie: `.out`-Dateien, `.log`-Dateien und `.err`-Dateien. Nur `.err`-Dateien werden komprimiert und aus dem Dateisystem entfernt, während `.log`- und `.out`-Protokolldateien im Dateisystem verbleiben. Um sicherzustellen, dass diese Ausgabedateien verwaltbar bleiben und der Cluster stabil bleibt, können Sie die Protokollrotation in `log4j.properties` so konfigurieren, dass eine maximale Anzahl von Dateien festgelegt und deren Größe begrenzt wird.

Amazon EMR 5.30.0 und höher

Ab Amazon EMR 5.30.0 verwendet Flink das Logging-Framework `log4j2` mit dem Namen der Konfigurationsklassifikation `flink-log4j`. Die folgende Beispielkonfiguration demonstriert das `log4j2`-Format.

```
[
  {
    "Classification": "flink-log4j",
    "Properties": {
      "appender.main.name": "MainAppender",
      "appender.main.type": "RollingFile",
      "appender.main.append" : "false",
```

```

    "appender.main.fileName" : "${sys:log.file}",
    "appender.main.filePattern" : "${sys:log.file}.%i",
    "appender.main.layout.type" : "PatternLayout",
    "appender.main.layout.pattern" : "%d{yyyy-MM-dd HH:mm:ss,SSS} %-5p %-60c %x - %m
%n",
    "appender.main.policies.type" : "Policies",
    "appender.main.policies.size.type" : "SizeBasedTriggeringPolicy",
    "appender.main.policies.size.size" : "100MB",
    "appender.main.strategy.type" : "DefaultRolloverStrategy",
    "appender.main.strategy.max" : "10"
  },
}
]

```

## Amazon-EMR-Versionen 5.29.0 und früher

Mit den Amazon-EMR-Versionen 5.29.0 und früher verwendet Flink das Logging-Framework log4j. Die folgende Beispielkonfiguration veranschaulicht das log4j Format.

```

[
  {
    "Classification": "flink-log4j",
    "Properties": {
      "log4j.appender.file": "org.apache.log4j.RollingFileAppender",
      "log4j.appender.file.append":"true",
      # keep up to 4 files and each file size is limited to 100MB
      "log4j.appender.file.MaxFileSize":"100MB",
      "log4j.appender.file.MaxBackupIndex":4,
      "log4j.appender.file.layout":"org.apache.log4j.PatternLayout",
      "log4j.appender.file.layout.ConversionPattern":"%d{yyyy-MM-dd HH:mm:ss,SSS} %-5p
%-60c %x - %m%n"
    },
  }
]

```

## Flink für die Ausführung mit Java 11 konfigurieren

Amazon-EMR-Versionen 6.12.0 und höher bieten Java-11-Laufzeitunterstützung für Flink. In den folgenden Abschnitten wird beschrieben, wie der Cluster so konfiguriert wird, dass er Java-11-Laufzeitunterstützung für Flink bereitstellt.

### Themen



- [Konfigurieren Sie Flink für Java 11, wenn Sie einen Cluster erstellen](#)
- [Flink für Java 11 auf einem laufenden Cluster konfigurieren](#)
- [Bestätigen Sie die Java-Laufzeit für Flink auf einem laufenden Cluster](#)

## Konfigurieren Sie Flink für Java 11, wenn Sie einen Cluster erstellen

Führen Sie die folgenden Schritte aus, um einen EMR-Cluster mit Flink und Java-11-Laufzeit zu erstellen. Die Konfigurationsdatei, in der Sie die Java-11-Laufzeitunterstützung hinzufügen, befindet sich `flink-conf.yaml`

### New console

Um einen Cluster mit Flink und Java-11-Laufzeit in der neuen Konsole zu erstellen

1. Melden Sie sich bei der AWS Management Console an und öffnen Sie die Amazon EMR-Konsole unter <https://console.aws.amazon.com/emr>.
2. Wählen Sie im Navigationsbereich unter EMR in EC2 die Option Cluster und dann Cluster erstellen aus.
3. Wählen Sie Amazon-EMR-Version 6.12.0 oder höher und installieren Sie die Flink-Anwendung. Wählen Sie alle anderen Anwendungen aus, die Sie auf Ihrem Cluster installieren möchten.
4. Fahren Sie mit der Einrichtung Ihres Clusters fort. Verwenden Sie im Bereich optionale Softwareeinstellungen die Standardoption Konfiguration eingeben und geben Sie die folgende Konfiguration ein:

```
[
  {
    "Classification": "flink-conf",
    "Properties": {
      "containerized.taskmanager.env.JAVA_HOME":"/usr/lib/jvm/jre-11",
      "containerized.master.env.JAVA_HOME":"/usr/lib/jvm/jre-11",
      "env.java.home":"/usr/lib/jvm/jre-11"
    }
  }
]
```

5. Fahren Sie mit der Einrichtung und dem Start Ihres Clusters fort.

## AWS CLI

Um einen Cluster mit Flink und Java-11-Laufzeit in der neuen Konsole zu erstellen

1. Erstellen Sie eine Konfigurationsdatei `configurations.json`, die Flink für die Verwendung von Java 11 konfiguriert.

```
[
  {
    "Classification": "flink-conf",
    "Properties": {
      "containerized.taskmanager.env.JAVA_HOME":"/usr/lib/jvm/jre-11",
      "containerized.master.env.JAVA_HOME":"/usr/lib/jvm/jre-11",
      "env.java.home":"/usr/lib/jvm/jre-11"
    }
  }
]
```

2. Erstellen Sie aus dem AWS CLI einen neuen EMR-Cluster mit Amazon EMR Version 6.12.0 oder höher und installieren Sie die Flink-Anwendung, wie im folgenden Beispiel gezeigt:

```
aws emr create-cluster --release-label emr-6.12.0 \
--applications Name=Flink \
--configurations file://./configurations.json \
--region us-east-1 \
--log-uri s3://myLogUri \
--instance-type m5.xlarge \
--instance-count 2 \
--service-role EMR_DefaultRole_V2 \
--ec2-attributes KeyName=YourKeyName,InstanceProfile=EMR_EC2_DefaultRole
```

## Flink für Java 11 auf einem laufenden Cluster konfigurieren

Gehen Sie wie folgt vor, um einen laufenden EMR-Cluster mit Flink und Java 11-Laufzeit zu aktualisieren. Die Konfigurationsdatei, in der Sie die Java-11-Laufzeitunterstützung hinzufügen, befindet sich `flink-conf.yaml`

## New console

Um einen laufenden Cluster mit Flink und Java-11-Laufzeit in der neuen Konsole zu aktualisieren

1. Melden Sie sich bei der AWS Management Console an und öffnen Sie die Amazon EMR-Konsole unter <https://console.aws.amazon.com/emr>.
2. Wählen Sie im Navigationsbereich unter EMR in EC2 die Option Cluster und dann den Cluster aus, den Sie aktualisieren möchten.

### Note

Der Cluster muss Amazon-EMR-Version 6.12.0 oder höher verwenden, um Java 11 zu unterstützen.

3. Wählen Sie die Registerkarte Konfigurationen aus.
4. Wählen Sie im Abschnitt Instance-Gruppenkonfigurationen die Instance-Gruppe Running aus, die Sie aktualisieren möchten, und wählen Sie dann Neukonfiguration aus dem Menü mit den Listenaktionen aus.
5. Konfigurieren Sie die Instance-Gruppe mit der Option Attribute bearbeiten wie folgt neu. Wählen Sie nach jeder Konfiguration Neue Konfiguration hinzufügen aus.

Klassifizierung	Eigenschaft	Wert
flink-conf	containerized.taskmanager.env.JAVA_HOME	/usr/lib/jvm/jre-11
flink-conf	containerized.master.env.JAVA_HOME	/usr/lib/jvm/jre-11
flink-conf	env.java.home	/usr/lib/jvm/jre-11

6. Wählen Sie Änderungen speichern aus um die Konfigurationen hinzuzufügen. .

## AWS CLI

Um einen laufenden Cluster für die Verwendung von Flink und Java-11-Laufzeit über die CLI zu aktualisieren

Verwenden Sie den Befehl `modify-instance-groups`, um eine neue Konfiguration für eine Instance-Gruppe in einem laufenden Cluster anzugeben.

1. Erstellen Sie eine Konfigurationsdatei `configurations.json`, die Flink für die Verwendung von Java 11 konfiguriert. Ersetzen Sie im folgenden Beispiel `ig-1xxxxxxx9` durch die ID für die Instance-Gruppe, die Sie neu konfigurieren möchten. Speichern Sie die Datei im selben Verzeichnis, in dem Sie den `modify-instance-groups`-Befehl ausführen werden.

```
[
  {
    "InstanceGroupId": "ig-1xxxxxxx9",
    "Configurations": [
      {
        "Classification": "flink-conf",
        "Properties": {
          "containerized.taskmanager.env.JAVA_HOME": "/usr/lib/jvm/jre-11",
          "containerized.master.env.JAVA_HOME": "/usr/lib/jvm/jre-11",
          "env.java.home": "/usr/lib/jvm/jre-11"
        },
        "Configurations": []
      }
    ]
  }
]
```

2. Führen Sie von der AWS CLI aus den folgenden Befehl aus. Ersetzen Sie die ID für die Instance-Gruppe, die Sie neu konfigurieren möchten:

```
aws emr modify-instance-groups --cluster-id j-2AL4XXXXXX5T9 \
--instance-groups file://configurations.json
```

## Bestätigen Sie die Java-Laufzeit für Flink auf einem laufenden Cluster

Um die Java-Laufzeit für einen laufenden Cluster zu ermitteln, melden Sie sich mit SSH beim Primärknoten an, wie unter [SSH mit dem Primärknoten verbinden](#) beschrieben. Führen Sie anschließend den folgenden Befehl aus:

```
ps -ef | grep flink
```

Der `ps`-Befehl mit der `-ef`-Option listet alle laufenden Prozesse auf dem System auf. Sie können diese Ausgabe mit `grep` filtern, um Erwähnungen der Zeichenfolge `flink` zu finden. Überprüfen Sie die Ausgabe für den Java-Laufzeitumgebung (JRE)-Wert, `jre-XX`. In der folgenden Ausgabe wird `jre-11` angegeben, dass Java 11 zur Laufzeit von Flink abgerufen wird.

```
flink    19130      1  0 09:17 ?          00:00:15 /usr/lib/jvm/jre-11/bin/
java -Djava.io.tmpdir=/mnt/tmp -Dlog.file=/usr/lib/flink/log/flink-flink-
historyserver-0-ip-172-31-32-127.log -Dlog4j.configuration=file:/usr/lib/flink/conf/
log4j.properties -Dlog4j.configurationFile=file:/usr/lib/flink/conf/log4j.properties
-Dlogback.configurationFile=file:/usr/lib/flink/conf/logback.xml -classpath /usr/lib/
flink/lib/flink-cep-1.17.0.jar:/usr/lib/flink/lib/flink-connector-files-1.17.0.jar:/
usr/lib/flink/lib/flink-csv-1.17.0.jar:/usr/lib/flink/lib/flink-json-1.17.0.jar:/usr/
lib/flink/lib/flink-scala_2.12-1.17.0.jar:/usr/lib/flink/lib/flink-table-api-java-
uber-1.17.0.jar:/usr/lib/flink/lib/flink-table-api-scala-bridge_2.12-1.17.0.
```

Alternativ können Sie [sich mit SSH am Primärknoten anmelden](#) und eine Flink YARN-Sitzung mit dem Befehl `flink-yarn-session -d` starten. Die Ausgabe zeigt die Java Virtual Machine (JVM) für Flink, `java-11-amazon-corretto` im folgenden Beispiel:

```
2023-05-29 10:38:14,129 INFO  org.apache.flink.configuration.GlobalConfiguration
[] - Loading configuration property: containerized.master.env.JAVA_HOME, /usr/lib/
jvm/java-11-amazon-corretto.x86_64
```

## Arbeiten mit Flink-Jobs in Amazon EMR

Es gibt mehrere Möglichkeiten, mit Flink in Amazon EMR zu interagieren: über die Konsole, die Flink-Schnittstelle in der Resource Manager Tracking-Benutzeroberfläche und in der Befehlszeile. Mit jedem dieser Programme können Sie eine JAR-Datei an eine Flink-Anwendung senden. Sobald Sie eine JAR-Datei eingereicht haben, wird sie zu einem Auftrag, der von Flink verwaltet wird JobManager. Die JobManager befindet sich auf dem YARN-Knoten, der den Application Master-Daemon der Flink-Sitzung hostet.

Sie können eine Flink-Anwendung als YARN-Auftrag auf einem Cluster mit langer Laufzeit oder auf einem vorübergehenden Cluster ausführen. In einem Cluster mit langer Laufzeit können Sie mehrere Flink-Aufträge an einen Flink-Cluster senden, der auf Amazon EMR ausgeführt wird. Wenn Sie einen Flink-Auftrag auf einem vorübergehenden Cluster ausführen, existiert Ihr Amazon-EMR-Cluster nur für die Zeit, die zum Ausführen der Flink-Anwendung benötigt wird, sodass Ihnen nur die verbrauchten Ressourcen und die Zeit in Rechnung gestellt werden. Sie können einen Flink-Auftrag mit der Amazon-EMR-AddSteps-API-Operation als Schrittargument für die RunJobFlow-Operation und über die Befehle AWS CLI, `add-steps` oder `create-cluster` übermitteln.

## Eine Flink-YARN-Anwendung als Schritt auf einem Cluster mit langer Laufzeit starten

Um eine Flink-Anwendung zu starten, an die mehrere Clients über YARN-API-Operationen Arbeit einreichen können, müssen Sie entweder einen Cluster erstellen oder eine Flink-Anwendung zu einem vorhandenen Cluster hinzufügen. Eine Anleitung zur Erstellung eines neuen Clusters finden Sie unter [Erstellen eines Clusters mit Flink](#). Um eine YARN-Sitzung auf einem vorhandenen Cluster zu starten, führen Sie die folgenden Schritte über die Konsole, AWS CLI oder das Java-SDK aus.

### Note

Der Befehl `flink-yarn-session` wurde in Amazon-EMR-Version 5.5.0 als Wrapper für das Skript `yarn-session.sh` zur Vereinfachung der Ausführung hinzugefügt. Wenn Sie eine frühere Version von Amazon EMR verwenden, ersetzen Sie `bash -c "/usr/lib/flink/bin/yarn-session.sh -d"` mit Argumente in der Konsole oder `Args` im AWS CLI-Befehl.

So senden Sie einen Flink-Auftrag auf einem vorhandenen Cluster von der Konsole aus

Senden Sie die Flink-Sitzung mit dem Befehl `flink-yarn-session` in einem vorhandenen Cluster.

1. Öffnen Sie die Amazon-EMR-Konsole unter <https://console.aws.amazon.com/emr>.
2. Wählen Sie in der Cluster-Liste den Cluster aus, den Sie zuvor gestartet haben.
3. Wählen Sie auf der Cluster-Detailseite **Steps (Schritte)** und **Add Step (Schritt hinzufügen)** aus.
4. Befolgen Sie die folgenden Richtlinien, um die Parameter einzugeben, und wählen Sie dann **Hinzufügen** aus.

Parameter	Beschreibung
Step type (Schritttyp)	Custom JAR
Name	Ein Name zur Identifizierung des Schritts. Zum Beispiel <i>&lt;example-flink-step-name&gt;</i> .
Jar location (Jar-Ort)	command-runner.jar
Argumente	Der Befehl <code>flink-yarn-session</code> mit für Ihre Anwendung geeigneten Argumenten. Beispielsweise startet <i>flink-yarn-session -d</i> eine Flink-Sitzung in Ihrem YARN-Cluster in einem getrennten Zustand (-d). Weitere Informationen finden Sie unter <a href="#">YARN-Einrichtung</a> in der aktuellen Flink-Dokumentation für Argumentdetails.

So senden Sie einen Flink-Auftrag auf einem vorhandenen Cluster mit der AWS CLI aus

- Verwenden Sie den `add-steps`-Befehl, um einem Cluster mit langer Laufzeit einen Flink-Auftrag hinzuzufügen. Der folgende Beispielbefehl gibt `Args="flink-yarn-session", "-d"` an, sodass eine Flink-Sitzung innerhalb Ihres YARN-Clusters in einem getrennten Zustand (-d) gestartet werden soll. Weitere Informationen finden Sie unter [YARN-Einrichtung](#) in der aktuellen Flink-Dokumentation für Argumentdetails.

```
aws emr add-steps --cluster-id <j-XXXXXXXX> --steps Type=CUSTOM_JAR,Name=<example-flink-step-name>,Jar=command-runner.jar,Args="flink-yarn-session","-d"
```

## Senden Sie Ihre Arbeit an eine bestehende Flink-Anwendung auf einem Cluster mit langer Laufzeit

Wenn Sie bereits eine bestehende Flink-Anwendung auf einem Cluster mit langer Laufzeit haben, können Sie die Flink-Anwendungs-ID des Clusters angeben, um Arbeit an diesen zu senden. Um die

Anwendungs-ID zu erhalten, führen Sie `yarn application -list` auf der AWS CLI oder über die [YarnClient](#) -API-Operation aus:

```
$ yarn application -list
16/09/07 19:32:13 INFO client.RMPProxy: Connecting to ResourceManager at
ip-10-181-83-19.ec2.internal/10.181.83.19:8032
Total number of applications (application-types: [] and states: [SUBMITTED, ACCEPTED,
RUNNING]):1
Application-Id      Application-Name      Application-Type      User      Queue      State
Final-State        Progress      Tracking-URL
application_1473169569237_0002      Flink session with 14 TaskManagers (detached)
  Apache Flink      hadoop      default      RUNNING      UNDEFINED
100% http://ip-10-136-154-194.ec2.internal:33089
```

Die Anwendungs-ID für diese Flink-Sitzung lautet `application_1473169569237_0002`, mit der Sie Arbeit aus dem AWS CLI oder einem SDK an die Anwendung senden können.

### Example SDK für Java

```
List<StepConfig> stepConfigs = new ArrayList<StepConfig>();

HadoopJarStepConfig flinkWordCountConf = new HadoopJarStepConfig()
    .withJar("command-runner.jar")
    .withArgs("flink", "run", "-m", "yarn-cluster", "-yid",
"application_1473169569237_0002", "-yn", "2", "/usr/lib/flink/examples/streaming/
WordCount.jar",
    "--input", "s3://myBucket/pg11.txt", "--output", "s3://myBucket/alice2/");

StepConfig flinkRunWordCount = new StepConfig()
    .withName("Flink add a wordcount step")
    .withActionOnFailure("CONTINUE")
    .withHadoopJarStep(flinkWordCountConf);

stepConfigs.add(flinkRunWordCount);

AddJobFlowStepsResult res = emr.addJobFlowSteps(new AddJobFlowStepsRequest()
    .withJobFlowId("myClusterId")
    .withSteps(stepConfigs));
```

### Example AWS CLI

```
aws emr add-steps --cluster-id <j-XXXXXXXX> \
```



```
--steps Type=CUSTOM_JAR,Name=Flink_Submit_To_Long_Running,Jar=command-runner.jar,\
Args="flink","run","-m","yarn-cluster","-yid","application_1473169569237_0002",\
"/usr/lib/flink/examples/streaming/WordCount.jar",\
"--input","s3://myBucket/pg11.txt","--output","s3://myBucket/alice2/" \
--region <region-code>
```

## Senden eines kurzlebigen Flink-Auftrags

Die folgenden Beispiele starten einen vorübergehenden Cluster, der einen Flink-Auftrag ausführt und dann nach Abschluss beendet wird.

### Example SDK für Java

```
import java.util.ArrayList;
import java.util.List;
import com.amazonaws.AmazonClientException;
import com.amazonaws.auth.AWSCredentials;
import com.amazonaws.auth.AWSStaticCredentialsProvider;
import com.amazonaws.auth.profile.ProfileCredentialsProvider;
import com.amazonaws.services.elasticmapreduce.AmazonElasticMapReduce;
import com.amazonaws.services.elasticmapreduce.AmazonElasticMapReduceClientBuilder;
import com.amazonaws.services.elasticmapreduce.model.*;

public class Main_test {

    public static void main(String[] args) {
        AWSCredentials credentials_profile = null;
        try {
            credentials_profile = new ProfileCredentialsProvider("default").getCredentials();
        } catch (Exception e) {
            throw new AmazonClientException(
                "Cannot load credentials from .aws/credentials file. " +
                "Make sure that the credentials file exists and the profile name is specified\n"
                "within it.",
                e);
        }

        AmazonElasticMapReduce emr = AmazonElasticMapReduceClientBuilder.standard()
            .withCredentials(new AWSStaticCredentialsProvider(credentials_profile))
            .withRegion(Regions.US_WEST_1)
            .build();

        List<StepConfig> stepConfigs = new ArrayList<StepConfig>();
```

```
HadoopJarStepConfig flinkWordCountConf = new HadoopJarStepConfig()
    .withJar("command-runner.jar")
    .withArgs("bash", "-c", "flink", "run", "-m", "yarn-cluster", "-yn", "2",
        "/usr/lib/flink/examples/streaming/WordCount.jar", "--input", "s3://path/to/
input-file.txt", "--output",
        "s3://path/to/output/");

StepConfig flinkRunWordCountStep = new StepConfig()
    .withName("Flink add a wordcount step and terminate")
    .withActionOnFailure("CONTINUE")
    .withHadoopJarStep(flinkWordCountConf);

stepConfigs.add(flinkRunWordCountStep);

Application flink = new Application().withName("Flink");

RunJobFlowRequest request = new RunJobFlowRequest()
    .withName("flink-transient")
    .withReleaseLabel("emr-5.20.0")
    .withApplications(flink)
    .withServiceRole("EMR_DefaultRole")
    .withJobFlowRole("EMR_EC2_DefaultRole")
    .withLogUri("s3://path/to/my/logfiles")
    .withInstances(new JobFlowInstancesConfig()
        .withEc2KeyName("myEc2Key")
        .withEc2SubnetId("subnet-12ab3c45")
        .withInstanceCount(3)
        .withKeepJobFlowAliveWhenNoSteps(false)
        .withMasterInstanceType("m4.large")
        .withSlaveInstanceType("m4.large"))
    .withSteps(stepConfigs);

RunJobFlowResult result = emr.runJobFlow(request);
System.out.println("The cluster ID is " + result.toString());

}

}
```

## Example AWS CLI

Verwenden Sie den Unterbefehl `create-cluster`, um einen kurzlebigen Cluster zu erstellen, der beendet wird, wenn der Flink-Auftrag abgeschlossen ist:

```
aws emr create-cluster --release-label emr-5.2.1 \  
--name "Flink_Transient" \  
--applications Name=Flink \  
--configurations file://./configurations.json \  
--region us-east-1 \  
--log-uri s3://myLogUri \  
--auto-terminate \  
--instance-type m5.xlarge \  
--instance-count 2 \  
--service-role EMR_DefaultRole_V2 \  
--ec2-attributes KeyName=<YourKeyName>,InstanceProfile=EMR_EC2_DefaultRole \  
--steps Type=CUSTOM_JAR,Jar=command-runner.jar,Name=Flink_Long_Running_Session,\  
Args="bash","-c","\\"flink run -m yarn-cluster /usr/lib/flink/examples/streaming/  
WordCount.jar  
--input s3://myBucket/pg11.txt --output s3://myBucket/alice/""
```

## Verwenden der Scala-Shell

Die Flink-Scala-Shell für EMR-Cluster ist so konfiguriert, dass nur neue YARN Sitzungen starten. Sie können die Scala-Shell wie im Folgenden beschrieben verwenden.

Den Flink Scala-Shell auf dem Primärknoten verwenden

1. Melden Sie sich mit SSH beim Primärknoten an, wie unter [SSH mit dem Primärknoten verbinden](#) beschrieben.
2. Geben Sie Folgendes ein, um eine Shell zu starten:

In Amazon-EMR-Version 5.5.0 und höher können Sie den folgenden Befehl verwenden, um einen Yarn-Cluster für die Scala-Shell mit einem TaskManager zu starten.

```
% flink-scala-shell yarn 1
```

In früheren Versionen von Amazon EMR verwenden Sie:

```
% /usr/lib/flink/bin/start-scala-shell.sh yarn 1
```

Damit wird die Flink Scala-Shell gestartet, sodass Sie Flink interaktiv nutzen können. Wie bei anderen Schnittstellen und Optionen können Sie im Beispiel verwendeten Optionswert `-n` basierend auf der Anzahl der Aufgaben, die Sie in der Shell ausführen möchten, skalieren.

Weitere Informationen finden Sie unter [Scala REPL](#) in der offiziellen Apache-Flink-Dokumentation.

## Suchen der Flink Webschnittstelle

Der Application Master, der zur Flink-Anwendung gehört, hostet das Flink-Webinterface. Dies ist eine alternative Möglichkeit, ein JAR als Auftrag einzureichen oder den aktuellen Status anderer Aufträge einzusehen. Die Flink-Webschnittstelle ist aktiv, solange eine Flink-Sitzung ausgeführt wird. Wenn bereits ein lang laufender YARN-Auftrag aktiv ist, können Sie die Anweisungen im Thema [Verbindung zum Primärknoten mit SSH](#) im Verwaltungshandbuch für Amazon EMR befolgen, um sich mit dem YARN ResourceManager zu verbinden. Wenn Sie beispielsweise einen SSH-Tunnel eingerichtet und in Ihrem Browser einen Proxy aktiviert haben, wählen Sie die ResourceManager-Verbindung unter Verbindungen auf der Cluster-Detailseite Ihres EMR-Clusters aus.

Cluster: Development Cluster Waiting Cluster ready after last step completed.

Connections:



[Resource Manager ... \(View All\)](#)

Nachdem Sie den ResourceManager gefunden haben, wählen Sie die YARN-Anwendung aus, die eine Flink-Sitzung hostet. Wählen Sie den Link unter der Spalte Tracking UI aus.

Lo

## All Applications

Containers Running	Memory Used	Memory Total	Memory Reserved	VCores Used	VCores Total	VCores Reserved	Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes	Unhealed Nodes
2	2 GB	11.25 GB	0 B	2	8	0	1	0	0	0	0

Scheduling Resource Type	Minimum Allocation	Maximum Allocation
MEMORY	<memory:32, vCores:1>	<memory:11520, vCores:8>

Name	Application Type	Queue	StartTime	FinishTime	State	FinalStatus	Progress	Tracking UI
Flink session with 1 TaskManagers (detached)	Apache Flink	default	Mon Oct 10 14:42:47 -0700 2016	N/A	RUNNING	UNDEFINED	<div style="width: 100%; height: 10px; background-color: #ccc;"></div>	<a href="#">ApplicationMaster</a>



In der Flink-Webschnittstelle können Sie die Konfiguration ansehen, Ihre eigene benutzerdefinierte JAR-Datei als Auftrag senden oder in Bearbeitung befindliche Aufträge überwachen.

**Apache Flink Dashboard** Overview Version: 1.1.1

Total Jobs	Count
Running	0
Finished	0
Canceled	0
Failed	0

Start Time	End Time	Duration	Job Name	Job ID	Tasks	Status
Running Jobs						
Completed Jobs						

## Flink Autoscaler

### Übersicht

Amazon-EMR-Versionen 6.15.0 und höher unterstützen Flink Autoscaler. Die Auftrag-Autoscaler-Funktion sammelt Metriken von laufenden Flink-Streaming-Aufträgen und skaliert automatisch die einzelnen Scheitelpunkte der Aufträge. Dadurch wird der Gegendruck reduziert und das von Ihnen festgelegte Nutzungsziel wird erfüllt.

Weitere Informationen finden Sie im Abschnitt [Autoscaler](#) in der Dokumentation zu Apache Flink Kubernetes Operator.

### Überlegungen

- Flink Autoscaler wird mit Amazon EMR 6.15.0 und höher unterstützt.
- Flink Autoscaler wird nur für Streaming-Aufträge unterstützt.
- Nur der adaptive Scheduler wird unterstützt. Der Standard-Scheduler wird nicht unterstützt.
- Wir empfehlen Ihnen, die Cluster-Skalierung zu aktivieren, um eine dynamische Ressourcenbereitstellung zu ermöglichen. Amazon EMR Managed Scaling wird bevorzugt, weil die Metrikauswertung alle 5 bis 10 Sekunden erfolgt. In diesem Intervall kann sich Ihr Cluster leichter an die Änderung der erforderlichen Cluster-Ressourcen anpassen.

## Autoscaler aktivieren

Gehen Sie wie folgt vor, um den Flink Autoscaler zu aktivieren, wenn Sie einen Amazon EMR in einem EC2-Cluster erstellen.

1. Erstellen Sie in der Amazon-EMR-Konsole einen neuen EMR-Cluster:
  - a. Wählen Sie Amazon EMR Version `emr-6.15.0` oder höher aus. Wählen Sie das Flink-Anwendungspaket und alle anderen Anwendungen aus, die Sie möglicherweise in Ihren Cluster aufnehmen möchten.

**Name and applications** [Info](#)

Name

Flink cluster

**Amazon EMR release** [Info](#)

A release contains a set of applications which can be installed on your cluster.

emr-6.15.0

**Application bundle**

Spark Interactive	Core Hadoop	Flink	HBase	Presto	Trino	Custom

Flink 1.17.1

HCatalog 3.1.3

Hue 4.11.0

Livy 0.7.1

Phoenix 5.1.3

Spark 3.4.1

Tez 0.10.2

ZooKeeper 3.5.10

Ganglia 3.7.2

Hadoop 3.3.6

JupyterEnterpriseGateway 2.6.0

MXNet 1.9.1

Pig 0.17.0

Sqoop 1.4.7

Trino 426

HBase 2.4.17

Hive 3.1.3

JupyterHub 1.5.0

Oozie 5.2.1

Presto 0.283

TensorFlow 2.11.0

Zeppelin 0.10.1

- b. Wählen Sie unter Option Cluster scaling and provisioning (Cluster-Skalierung und -Bereitstellung) Use EMR-managed scaling (EMR-verwaltete Skalierung verwenden) aus.

## Cluster scaling and provisioning [Info](#)

Set up scaling and provisioning configurations for the core and task node groups for your cluster.

Choose an option

Set cluster size manually  
Use this option if you know your workload patterns in advance.

Use EMR-managed scaling  
Monitor key workload metrics so that EMR can optimize the cluster size and resource utilization.

Use custom automatic scaling  
To programmatically scale core and task nodes, create custom automatic scaling policies.

2. Geben Sie im Abschnitt `Software settings` (Softwareeinstellungen) die folgende Konfiguration ein, um Flink Autoscaler zu aktivieren. Stellen Sie für Testszzenarien das Entscheidungsintervall, das Metrikfensterintervall und das Stabilisierungsintervall auf einen niedrigeren Wert ein, sodass der Auftrag sofort eine Skalierungsentscheidung trifft, um die Überprüfung zu vereinfachen.

```
[
  {
    "Classification": "flink-conf",
    "Properties": {
      "job.autoscaler.enabled": "true",
      "jobmanager.scheduler": "adaptive",
      "job.autoscaler.stabilization.interval": "60s",
      "job.autoscaler.metrics.window": "60s",
      "job.autoscaler.decision.interval": "10s",
      "job.autoscaler.debug.logs.interval": "60s"
    }
  }
]
```

3. Wählen oder konfigurieren Sie alle anderen Einstellungen nach Ihren Wünschen und erstellen Sie den Flink-Autoscaler-fähigen Cluster.

## Autoscaler-Konfigurationen

In diesem Abschnitt werden die meisten Konfigurationen behandelt, die Sie je nach Ihren spezifischen Anforderungen ändern können.

**Note**

Bei zeitbasierten Konfigurationen wie `time-`, `interval-` und `window-`Einstellungen ist die Standardeinheit, wenn keine Einheit angegeben ist, Millisekunden. Ein Wert von 30 ohne Suffix entspricht also 30 Millisekunden. Geben Sie für andere Zeiteinheiten das entsprechende Suffix `s` für Sekunden, `m` für Minuten oder `h` für Stunden an.

## Themen

- [Autoscaler-Loop-Konfigurationen](#)
- [Aggregation von Metriken und Verlaufskonfigurationen](#)
- [Konfigurationen auf Ebene des Auftrags Scheitelpunkts](#)
- [Konfigurationen für die Backlog-Verarbeitung](#)
- [Vorgangskonfigurationen skalieren](#)

## Autoscaler-Loop-Konfigurationen

Autoscaler ruft die Metriken auf Ebene des Auftrags Scheitelpunkts alle paar konfigurierbare Zeitintervalle ab, wandelt sie in umsetzbare Maßeinheiten um, schätzt die Parallelität neuer Auftrags Scheitelpunkte und empfiehlt sie dem Auftrags-Scheduler. Metriken werden erst nach dem Neustart des Auftrags und dem Cluster-Stabilisierungsintervall erfasst.

Konfigurationsschlüssel	Standardwert	Beschreibung	Beispielwerte
<code>job.autoscaler.enabled</code>	<code>false</code>	Aktivieren Sie die automatische Skalierung auf Ihrem Flink-Cluster.	<code>true</code> , <code>false</code>
<code>job.autoscaler.decision.interval</code>	<code>60s</code>	Autoscaler-Entscheidungsintervall.	30 (Standardeinheit ist Millisekunden), 5m, 1h
<code>job.autoscaler.restart.time</code>	<code>3m</code>	Die erwartete Neustartzeit wird verwendet, bis der Bediener	30 (Standardeinheit ist



Konfigurationsschlüssel	Standardwert	Beschreibung	Beispielwerte
		sie anhand des Verlaufs zuverlässig ermitteln kann.	Millisekunden), 5m, 1h
<code>job.autoscaler.stabilization.interval</code>	300s	Stabilisierungszeitraum, in dem keine neue Skalierung durchgeführt wird.	30 (Standard einheit ist Millisekunden), 5m, 1h
<code>job.autoscaler.debug.logs.interval</code>	300s	Intervall für Autoscaler-Debug-Protokolle.	30 (Standard einheit ist Millisekunden), 5m, 1h

## Aggregation von Metriken und Verlaufskonfigurationen

Autoscaler ruft die Metriken ab, aggregiert sie über ein zeitbasiertes gleitendes Fenster und wertet sie zu Skalierungsentscheidungen aus. Der Verlauf der Skalierungsentscheidungen für jeden Auftrags Scheitelpunkt wird verwendet, um die neue Parallelität abzuschätzen. Diese haben sowohl ein altersabhängiges Verfallsdatum als auch eine historische Größe (mindestens 1).

Konfigurationsschlüssel	Standardwert	Beschreibung	Beispielwerte
<code>job.autoscaler.metrics.window</code>	600s	Scaling metrics aggregation window size.	30 (Standard einheit ist Millisekunden), 5m, 1h
<code>job.autoscaler.history.max.count</code>	3	Maximale Anzahl vergangener Skalierungsentscheidungen, die pro Scheitelpunkt beibehalten werden sollen.	1 auf Integer.MAX_VALUE
<code>job.autoscaler.history.max.age</code>	24h	Mindestanzahl vergangener Skalierungsentscheidungen, die pro Scheitelpunkt beibehalten werden sollen.	30 (Standard einheit ist

Konfigurationsschlüssel	Standardwert	Beschreibung	Beispielwerte
			Millisekunden), 5m, 1h

## Konfigurationen auf Ebene des Auftragsscheitelpunkts

Die Parallelität jedes Auftragsscheitelpunkts wird auf der Grundlage der Zielauslastung geändert und durch die Min./Max.-Parallelitätsgrenzen begrenzt. Es wird nicht empfohlen, die Zielauslastung auf nahezu 100 % festzulegen (d. h. den Wert 1), und die Nutzungsgrenze dient als Puffer, um die zwischenzeitlichen Lastschwankungen zu bewältigen.

Konfigurationsschlüssel	Standardwert	Beschreibung	Beispielwerte
<code>job.autoscaler.target.utilization</code>	0.7	Nutzung des Zielscheitelpunkts.	0 - 1
<code>job.autoscaler.target.utilization.boundary</code>	0.4	Nutzung des Zielscheitelpunkts. Die Skalierung wird nicht durchgeführt, wenn die aktuelle Verarbeitungsrate innerhalb von $[\text{target\_rate} / (\text{target\_utilization} - \text{boundary})$ und $(\text{target\_rate} / (\text{target\_utilization} + \text{boundary}))$ liegt	0 - 1
<code>job.autoscaler.vertex.min-parallelism</code>	1	Die Mindestparallelität, die der Autoscaler verwenden kann.	0 - 200
<code>job.autoscaler.vertex.max-parallelism</code>	200	Die maximale Parallelität, die der Autoscaler verwenden kann. Der Autoscaler ignoriert dieses Limit, wenn es höher ist als die maximale Parallelität,	0 - 200

Konfigurationsschlüssel	Standardwert	Beschreibung	Beispielwerte
		die in der Flink-Konfiguration oder direkt für jeden Bediener konfiguriert wurde.	

## Konfigurationen für die Backlog-Verarbeitung

Der Auftrags Scheitelpunkt benötigt zusätzliche Ressourcen, um die ausstehenden Ereignisse oder Rückstände zu bewältigen, die sich während des Skalierungsvorgangs ansammeln.

Dies wird auch als die `catch-up`-Dauer bezeichnet. Wenn die Zeit für die Verarbeitung des Backlogs den konfigurierten `lag -threshold`-Wert überschreitet, steigt die Zielauslastung des Auftrags Scheitelpunkts auf den Höchstwert. Auf diese Weise können unnötige Skalierungsvorgänge während der Backlog-Verarbeitung vermieden werden.

Konfigurationsschlüssel	Standardwert	Beschreibung	Beispielwerte
<code>job.autoscaler.backlog-processing.lag-threshold</code>	5m	Verzögerungsschwellenwert, der unnötige Skalierungen verhindert und gleichzeitig die ausstehenden Nachrichten entfernt, die für die Verzögerung verantwortlich sind.	30 (Standard einheit ist Millisekunden), 5m, 1h
<code>job.autoscaler.catch-up.duration</code>	15m	Die Zieldauer für die vollständige Backlog-Verarbeitung nach einem Skalierungsvorgang. Auf 0 setzen, um die Backlog-basierte Skalierung zu deaktivieren.	30 (Standard einheit ist Millisekunden), 5m, 1h

## Vorgangskonfigurationen skalieren

Autoscaler führt die Herunterskalierung innerhalb einer Kulanzzzeit nicht unmittelbar nach einer Hochskalierung durch. Dadurch wird ein unnötiger Zyklus von Hoch- und

Herunterskalierungsvorgängen, die durch vorübergehende Lastschwankungen verursacht werden, vermieden.

Wir können das Verhältnis der Herunterskalierung verwenden, um die Parallelität schrittweise zu verringern und Ressourcen freizusetzen, um vorübergehende Lastspitzen auszugleichen. Das trägt auch dazu bei, unnötige kleinere Hochskalierungsvorgänge nach einer umfangreichen Herunterskalierung zu vermeiden.

Wir können einen ineffektiven Skalierungsvorgang auf der Grundlage vergangener Entscheidungen über die Skalierung von Auftragsscheitelpunkten erkennen, um weitere Änderungen der Parallelität zu verhindern.

Konfigurationsschlüssel	Standardwert	Beschreibung	Beispielwerte
<code>job.autoscaler.scale-up.grace-period</code>	1h	Dauer, in der ein Scheitelpunkt nicht herunterskaliert werden darf, nachdem er hochskaliert wurde.	30 (Standard einheit ist Millisekunden), 5m, 1h
<code>job.autoscaler.scale-down.max-factor</code>	0.6	Maximaler Herunterskalierungsfaktor. Ein Wert von 1 bedeutet, dass es keine Begrenzung der Herunterskalierung gibt. 0.6 bedeutet, dass der Auftrag nur mit 60 % der ursprünglichen Parallelität herunterskaliert werden kann.	0 - 1
<code>job.autoscaler.scale-up.max-factor</code>	100000.	Maximales Verhältnis der Hochskalierung. Der Wert 2.0 bedeutet, dass der Auftrag nur mit 200 % der aktuellen Parallelität hochskaliert werden kann.	0 - Integer.MAX_VALUE
<code>job.autoscaler.scaling.effectiveness.detection.enabled</code>	false	Ob ineffektive Skalierungsvorgänge erkannt werden sollen und ob der Autoscale	true, false

Konfigurationsschlüssel	Standardwert	Beschreibung	Beispielwerte
		r weitere Hochskalierungen blockieren kann.	

## Optimieren der Neustartzeiten von Aufträgen für die Aufgabenwiederherstellung und -skalierung

Wenn eine Aufgabe fehlschlägt oder wenn ein Skalierungsvorgang stattfindet, versucht Flink, die Aufgabe vom letzten abgeschlossenen Prüfpunkt aus erneut auszuführen. Die Ausführung des Neustartvorgangs kann eine Minute oder länger dauern, abhängig von der Größe des Prüfpunktzustands und der Anzahl der parallelen Aufgaben. Während des Neustarts können sich Backlog-Aufgaben für den Auftrag ansammeln. Es gibt jedoch einige Möglichkeiten, wie Flink die Geschwindigkeit der Wiederherstellung und des Neustarts von Ausführungsdiagrammen optimiert, um die Auftragsstabilität zu verbessern.

Auf dieser Seite werden einige der Möglichkeiten beschrieben, mit denen Amazon EMR Flink die Zeit für den Neustart des Auftrags während der Aufgabenwiederherstellung oder -skalierung verbessern kann.

### Themen

- [Aufgabenlokale Wiederherstellung](#)
- [Generischer protokollbasierter inkrementeller Prüfpunkt](#)
- [Differenzierte Wiederherstellung](#)
- [Kombinierter Neustartmechanismus im adaptiven Scheduler](#)

## Aufgabenlokale Wiederherstellung

### Note

Aufgabenlokale Wiederherstellung wird mit Amazon EMR 6.0.0 und höher unterstützt.

Mit Flink-Prüfpunkten erstellt jede Aufgabe einen Snapshot ihres Status, den Flink in verteilte Speicher wie Amazon S3 schreibt. Im Falle einer Wiederherstellung stellen die Aufgaben ihren Status

aus dem verteilten Speicher wieder her. Der verteilte Speicher bietet Fehlertoleranz und kann den Status während der Neuskalierung neu verteilen, da er für alle Knoten zugänglich ist.

Ein verteilter Remote-Speicher hat jedoch auch einen Nachteil: Alle Aufgaben müssen ihren Status von einem entfernten Standort aus über das Netzwerk lesen. Dies kann bei der Aufgabenwiederherstellung oder bei Skalierungsvorgängen zu langen Wiederherstellungszeiten für große Zustände führen.

Dieses Problem der langen Wiederherstellungszeit wird durch eine aufgabenlokale Wiederherstellung gelöst. Aufgaben schreiben ihren Status am Prüfpunkt in einen sekundären Speicher, der sich lokal zur Aufgabe befindet, z. B. auf eine lokale Festplatte. Sie speichern ihren Status auch im Primärspeicher oder in unserem Fall in Amazon S3. Während der Wiederherstellung plant der Scheduler die Aufgaben in demselben Task-Manager, in dem die Aufgaben zuvor ausgeführt wurden, sodass sie aus dem lokalen Statusspeicher wiederhergestellt werden können, anstatt sie aus dem Remote-Statusspeicher zu lesen. Weitere Informationen finden Sie unter [Aufgabenlokale Wiederherstellung](#) in der Apache-Flink-Dokumentation.

Unsere Benchmark-Tests mit Beispielaufträgen haben gezeigt, dass die Wiederherstellungszeit bei aktivierter aufgabenlokaler Wiederherstellung von Minuten auf wenige Sekunden reduziert wurde.

Um die aufgabenlokale Wiederherstellung zu aktivieren, legen Sie die folgenden Konfigurationen in Ihrer `flink-conf.yaml`-Datei fest. Geben Sie den Wert für das Prüfpunkt-Intervall in Millisekunden an.

```
state.backend.local-recovery: true
state.backend: hasmap or rocksdb
state.checkpoints.dir: s3://storage-location-bucket-path/checkpoint
execution.checkpointing.interval: 15000
```

## Generischer protokollbasierter inkrementeller Prüfpunkt

### Note

Generische protokollbasierte inkrementelle Prüfpunkte werden mit Amazon EMR 6.10.0 und höher unterstützt.

Generische protokollbasierte inkrementelle Prüfpunkte wurden in Flink 1.16 hinzugefügt, um die Geschwindigkeit von Prüfpunkten zu verbessern. Ein schnelleres Prüfpunktintervall führt

häufig zu einer Reduzierung des Wiederherstellungsaufwands, da weniger Ereignisse nach der Wiederherstellung erneut verarbeitet werden müssen. Weitere Informationen finden Sie im Apache-Flink-Blog unter [Verbesserung der Geschwindigkeit und Stabilität von Prüfpunkten mit generischen protokollbasierten inkrementellen Prüfpunkten](#).

Unsere Benchmark-Tests haben anhand von Beispielaufträgen gezeigt, dass sich die Prüfpunktzeit mit dem generischen protokollbasierten inkrementellen Prüfpunkt von Minuten auf wenige Sekunden reduziert hat.

Um generische protokollbasierte inkrementelle Prüfpunkte zu aktivieren, legen Sie die folgenden Konfigurationen in Ihrer Datei `flink-conf.yaml` fest. Geben Sie den Wert für das Prüfpunktintervall in Millisekunden an.

```
state.backend.changelog.enabled: true
state.backend.changelog.storage: filesystem
dstl.dfs.base-path: s3://bucket-path/changelog
state.backend.local-recovery: true
state.backend: rocksdb
state.checkpoints.dir: s3://bucket-path/checkpoint
execution.checkpointing.interval: 15000
```

## Differenzierte Wiederherstellung

### Note

Eine differenzierte Wiederherstellungsunterstützung für den Standard-Scheduler ist mit Amazon EMR 6.0.0 und höher verfügbar. Eine differenzierte Wiederherstellungsunterstützung im adaptiven Scheduler ist mit Amazon EMR 6.15.0 und höher verfügbar.

Wenn eine Aufgabe während der Ausführung fehlschlägt, setzt Flink das gesamte Ausführungsdiagramm zurück und löst eine vollständige Neuausführung ab dem letzten abgeschlossenen Prüfpunkt aus. Das ist teurer, als nur die fehlgeschlagenen Aufgaben erneut auszuführen. Bei einer differenzierten Wiederherstellung wird nur die mit der Pipeline verbundene Komponente der fehlgeschlagenen Aufgabe neu gestartet. Im folgenden Beispiel hat das Auftragsdiagramm 5 Scheitelpunkte (A bis E). Alle Verbindungen zwischen den Scheitelpunkten werden punktweise in Pipelines verlegt, und der Wert `parallelism.default` für den Auftrag ist auf 2 eingestellt.

```
A # B # C # D # E
```

In diesem Beispiel werden insgesamt 10 Aufgaben ausgeführt. Die erste Pipeline (a1 bis e1) wird in einem TaskManager (TM1) ausgeführt, während die zweite Pipeline (a2 bis e2) in einem anderen TaskManager (TM2) ausgeführt wird.

```
a1 # b1 # c1 # d1 # e1  
a2 # b2 # c2 # d2 # e2
```

Es gibt zwei Komponenten, die über eine Pipeline miteinander verbunden sind: a1 # e1 und a2 # e2. Wenn entweder TM1 oder TM2 fehlschlägt, wirkt sich der Fehler nur auf die 5 Aufgaben in der Pipeline aus, in denen der TaskManager ausgeführt wurde. Bei der Neustartstrategie wird nur die betroffene Pipeline-Komponente gestartet.

Eine differenzierte Wiederherstellung funktioniert nur mit perfekt parallelen Flink-Aufträgen. Sie wird nicht mit `keyBy()`- oder `reDistribute()`-Vorgängen unterstützt. Weitere Informationen finden Sie unter [FLIP-1: Fine Grained Recovery from Task Failures](#) (FLIP-1: Differenzierte Wiederherstellung nach Aufgabenfehlern) im Jira-Projekt Flink Improvement Proposal.

Um die differenzierte Wiederherstellung zu aktivieren, legen Sie die folgenden Konfigurationen in Ihrer `flink-conf.yaml`-Datei fest.

```
jobmanager.execution.failover-strategy: region  
restart-strategy: exponential-delay or fixed-delay
```

## Kombinierter Neustartmechanismus im adaptiven Scheduler

### Note

Der kombinierte Neustartmechanismus im adaptiven Scheduler wird mit Amazon EMR 6.15.0 und höher unterstützt.

Der adaptive Scheduler kann die Parallelität des Auftrags auf der Grundlage der verfügbaren Slots anpassen. Er reduziert automatisch die Parallelität, wenn nicht genügend Slots für die konfigurierte Auftragsparallelität verfügbar sind. Wenn neue Slots verfügbar werden, wird der Auftrag wieder auf die konfigurierte Auftragsparallelität hochskaliert. Ein adaptiver Scheduler vermeidet Ausfallzeiten



beim Auftrag, wenn nicht genügend Ressourcen verfügbar sind. Dies ist der unterstützte Scheduler für Flink Autoscaler. Aus diesen Gründen empfehlen wir den adaptiven Scheduler mit Amazon EMR Flink. Adaptive Scheduler können jedoch innerhalb kurzer Zeit mehrere Neustarts durchführen, und zwar einen Neustart für jede neu hinzugefügte Ressource. Dies könnte zu einem Leistungsabfall des Auftrags führen.

Mit Amazon EMR 6.15.0 und höher verfügt Flink über einen kombinierten Neustartmechanismus im adaptiven Scheduler, der ein Neustartfenster öffnet, wenn die erste Ressource hinzugefügt wird, und dann bis zum konfigurierten Fensterintervall von 1 Minute wartet. Er führt einen einzigen Neustart durch, wenn genügend Ressourcen zur Verfügung stehen, um den Auftrag mit konfigurierter Parallelität auszuführen, oder wenn das Intervall abgelaufen ist.

Unsere Benchmark-Tests haben anhand von Beispielaufträgen gezeigt, dass dieses Feature 10 % mehr Datensätze verarbeitet als das Standardverhalten, wenn Sie den adaptiven Scheduler und Flink Autoscaler verwenden.

Um den kombinierten Neustartmechanismus zu aktivieren, legen Sie die folgenden Konfigurationen in Ihrer Datei `flink-conf.yaml` fest.

```
jobmanager.adaptive-scheduler.combined-restart.enabled: true
jobmanager.adaptive-scheduler.combined-restart.window-interval: 1m
```

## Arbeiten mit Flink-Aufträgen von Zeppelin in Amazon EMR

### Einführung

Amazon-EMR-Versionen 6.10.0 und höher unterstützen die [Apache Zeppelin](#)-Integration mit Apache Flink. Sie können Flink-Aufträge interaktiv über Zeppelin-Notebooks einreichen. Mit dem Flink-Interpreter können Sie Flink-Abfragen ausführen, Flink-Streaming- und Batch-Aufträge definieren und die Ausgabe in Zeppelin-Notebooks visualisieren. Der Flink-Interpreter basiert auf der Flink-REST-API. Auf diese Weise können Sie von der Zeppelin-Umgebung aus auf Flink-Aufträge zugreifen und diese bearbeiten, um eine Datenverarbeitung und -analyse in Echtzeit durchzuführen.

In Flink Interpreter gibt es vier Unterinterpreter. Sie dienen unterschiedlichen Zwecken, befinden sich aber alle in der JVM und teilen sich dieselben vorkonfigurierten Einstiegspunkte zu Flink (`ExecutionEnvironment`, `StreamExecutionEnvironment`, `BatchTableEnvironment`, `StreamTableEnvironment`). Die Interpreter sind wie folgt:

- `%flink` – Erzeugt `ExecutionEnvironment`, `StreamExecutionEnvironment`, `BatchTableEnvironment`, `StreamTableEnvironment` und stellt eine Scala-Umgebung bereit
- `%flink.pyflink` – Stellt eine Python-Umgebung bereit
- `%flink.ssql` – Stellt eine Streaming-SQL-Umgebung bereit
- `%flink.bsql` – Stellt eine Batch-SQL-Umgebung bereit

## Voraussetzungen

- Die Zeppelin-Integration mit Flink wird für Cluster unterstützt, die mit Amazon EMR 6.10.0 und höher erstellt wurden.
- Um Webschnittstellen, die auf EMR-Clustern gehostet werden, wie für diese Schritte erforderlich, anzuzeigen, müssen Sie einen SSH-Tunnel konfigurieren, der eingehenden Zugriff ermöglicht. Weitere Informationen finden Sie unter [Konfigurieren von Proxy-Einstellungen, um auf dem Primärknoten gehostete Websites anzeigen zu lassen](#).

## Zeppelin-Flink auf einem EMR-Cluster konfigurieren

Gehen Sie wie folgt vor, um Apache Flink auf Apache Zeppelin für die Ausführung auf einem EMR-Cluster zu konfigurieren:


1. Erstellen Sie einen neuen Cluster von der Amazon-EMR-Konsole aus. Wählen Sie `emr-6.10.0` oder höher für die Amazon-EMR-Version aus. Wählen Sie dann, ob Sie Ihr Anwendungspaket mit der Option Benutzerdefiniert anpassen möchten. Nehmen Sie mindestens Flink, Hadoop und Zeppelin in Ihr Paket auf.


Amazon EMR release [Info](#)


A release contains a set of applications which can be installed on your cluster.


emr-6.10.0 ▼


Application bundle


Spark  


Core  
Hadoop  


HBase  


Presto  


Trino  


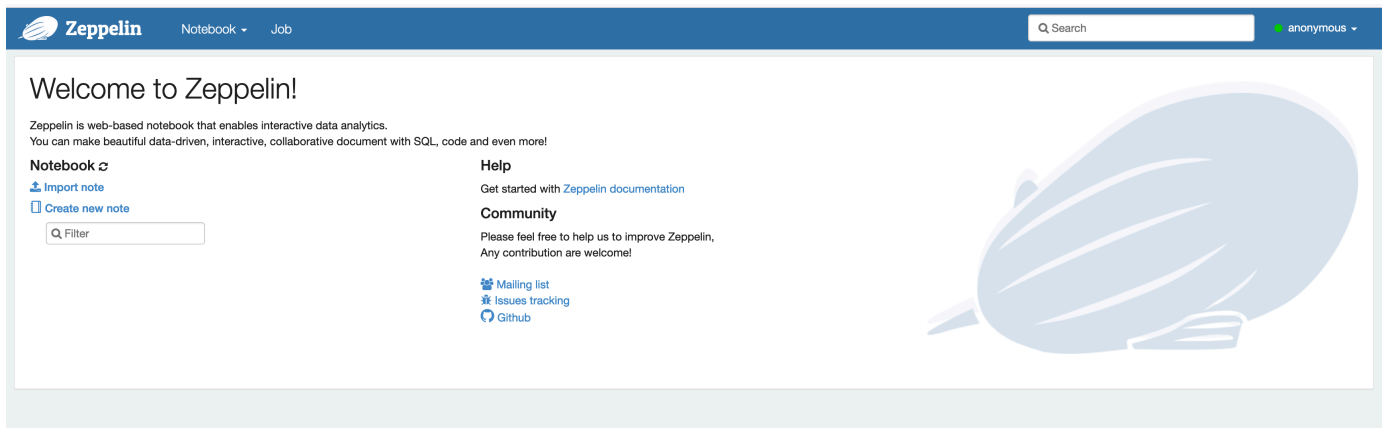
Custom  


▼ Customize your application bundle

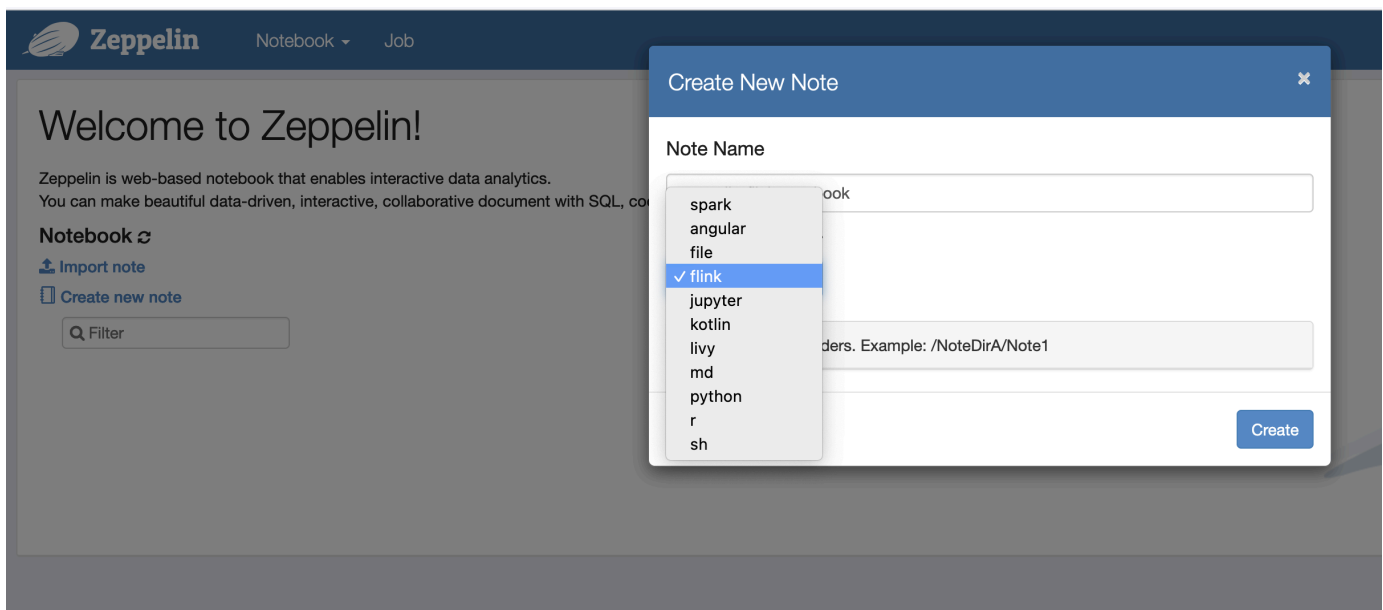
Applications included in bundle

<input checked="" type="checkbox"/> Flink 1.16.0	<input type="checkbox"/> Ganglia 3.7.2
<input type="checkbox"/> HBase 2.4.15	<input type="checkbox"/> HCatalog 3.1.3
<input checked="" type="checkbox"/> Hadoop 3.3.3	<input type="checkbox"/> Hive 3.1.3
<input type="checkbox"/> Hue 4.10.0	<input type="checkbox"/> JupyterEnterpriseGateway 2.6.0
<input type="checkbox"/> JupyterHub 1.5.0	<input type="checkbox"/> Livy 0.7.1
<input type="checkbox"/> MXNet 1.9.1	<input type="checkbox"/> Oozie 5.2.1
<input type="checkbox"/> Phoenix 5.1.2	<input type="checkbox"/> Pig 0.17.0
<input type="checkbox"/> Presto 0.278	<input type="checkbox"/> Spark 3.3.1
<input type="checkbox"/> Sqoop 1.4.7	<input type="checkbox"/> TensorFlow 2.11.0
<input type="checkbox"/> Tez 0.10.2	<input type="checkbox"/> Trino 403
<input checked="" type="checkbox"/> Zeppelin 0.10.1	<input type="checkbox"/> ZooKeeper 3.5.10

2. Erstellen Sie den Rest Ihres Clusters mit den Einstellungen, die Sie bevorzugen.
3. Sobald Ihr Cluster läuft, wählen Sie den Cluster in der Konsole aus, um seine Details anzuzeigen, und öffnen Sie die Registerkarte Anwendungen. Wählen Sie Zeppelin im Bereich Benutzeroberflächen für Anwendungen aus, um die Zeppelin-Weboberfläche zu öffnen. Stellen Sie sicher, dass Sie den Zugriff auf die Zeppelin-Weboberfläche mit einem SSH-Tunnel zum Primärknoten und einer Proxyverbindung eingerichtet haben, wie in [Voraussetzungen](#) beschrieben.



4. Jetzt können Sie eine neue Notiz in einem Zeppelin-Notebook mit Flink als Standardinterpreter erstellen.



5. In den folgenden Codebeispielen wird veranschaulicht, wie Flink-Jobs von einem Zeppelin-Notebook aus ausgeführt werden.

Führen Sie Flink-Aufträge mit Zeppelin-Flink auf einem EMR-Cluster aus

- Beispiel 1, Flink Scala

a) Batch- WordCount Beispiel (SCALA)

```
%flink

val data = benv.fromElements("hello world", "hello flink", "hello hadoop")
```

```
data.flatMap(line => line.split("\\s"))
    .map(w => (w, 1))
    .groupBy(0)
    .sum(1)
    .print()
```

## b) Streaming- WordCount Beispiel (SCALA)

```
%flink

val data = env.fromElements("hello world", "hello flink", "hello hadoop")
data.flatMap(line => line.split("\\s"))
    .map(w => (w, 1))
    .keyBy(0)
    .sum(1)
    .print

env.execute()
```

## • Beispiel 2, Flink Streaming SQL

```
%flink.sql
SET 'sql-client.execution.result-mode' = 'tableau';
SET 'table.dml-sync' = 'true';
SET 'execution.runtime-mode' = 'streaming';

create table dummy_table (
  id int,
  data string
) with (
  'connector' = 'filesystem',
  'path' = 's3://s3-bucket/dummy_table',
  'format' = 'csv'
```

```
);

INSERT INTO dummy_table SELECT * FROM (VALUES (1, 'Hello World'), (2, 'Hi'), (2, 'Hi'), (3, 'Hello'), (3, 'World'), (4, 'ADD'), (5, 'LINE'));

SELECT * FROM dummy_table;
```

The screenshot shows a PyFlink job execution interface. The top part displays the SQL code used for the job. Below the code, a table shows the results of the query. The table has two columns: 'id' and 'data'. The results are as follows:

id	data
3	Hello
3	World
3	Hello
3	World
4	ADD
4	ADD
5	LINE
5	LINE

At the bottom of the screenshot, there is a status bar that reads: "Took 20 sec. Last updated by anonymous at May 04 2023, 3:07:03 PM. (outdated)"

- Beispiel 3, Pyflink. Beachten Sie, dass Sie Ihre eigene Beispieldatei mit dem Namen `word.txt` in Ihren S3-Bucket hochladen müssen.

```
%flink.pyflink

import argparse
import logging
import sys

from pyflink.common import Row
from pyflink.table import (EnvironmentSettings, TableEnvironment, TableDescriptor,
                           Schema,
                           DataTypes, FormatDescriptor)
from pyflink.table.expressions import lit, col
from pyflink.table.udf import udtf

def word_count(input_path, output_path):
    t_env = TableEnvironment.create(EnvironmentSettings.in_streaming_mode())
    # write all the data to one file
    t_env.get_config().set("parallelism.default", "1")
```

```

# define the source
if input_path is not None:
    t_env.create_temporary_table(
        'source',
        TableDescriptor.for_connector('filesystem')
            .schema(Schema.new_builder()
                .column('word', DataTypes.STRING())
                .build())
            .option('path', input_path)
            .format('csv')
            .build())
    tab = t_env.from_path('source')
else:
    print("Executing word_count example with default input data set.")
    print("Use --input to specify file input.")
    tab = t_env.from_elements(map(lambda i: (i,), word_count_data),
        DataTypes.ROW([DataTypes.FIELD('line',
DataTypes.STRING())]))

# define the sink
if output_path is not None:
    t_env.create_temporary_table(
        'sink',
        TableDescriptor.for_connector('filesystem')
            .schema(Schema.new_builder()
                .column('word', DataTypes.STRING())
                .column('count', DataTypes.BIGINT())
                .build())
            .option('path', output_path)
            .format(FormatDescriptor.for_format('canal-json')
                .build())
            .build())
else:
    print("Printing result to stdout. Use --output to specify output path.")
    t_env.create_temporary_table(
        'sink',
        TableDescriptor.for_connector('print')
            .schema(Schema.new_builder()
                .column('word', DataTypes.STRING())
                .column('count', DataTypes.BIGINT())
                .build())
            .build())

```

```

@udtf(result_types=[DataTypes.STRING()])
def split(line: Row):
    for s in line[0].split():
        yield Row(s)

# compute word count
tab.flat_map(split).alias('word') \
    .group_by(col('word')) \
    .select(col('word'), lit(1).count) \
    .execute_insert('sink') \
    .wait()

logging.basicConfig(stream=sys.stdout, level=logging.INFO, format="%(message)s")

word_count("s3://s3_bucket/word.txt", "s3://s3_bucket/demo_output.txt")

```

1. Wählen Sie FLINK JOB in der Zeppelin-Benutzeroberfläche, um auf die Flink-Web-UI zuzugreifen und diese anzusehen.

**Batch WordCount**

FLINK JOB FINISHED

```

%flink

val data = benv.fromElements("hello world", "hello flink", "hello hadoop")
data.flatMap(line => line.split("\\s"))
    .map(w => (w, 1))
    .groupBy(0)
    .sum(1)
    .print()

data: org.apache.flink.api.scala.DataSet[String] = org.apache.flink.api.scala.DataSet@22fe7dd5
(flink,1)
(hadoop,1)
(hello,3)
(world,1)

Took 56 sec. Last updated by anonymous at May 04 2023, 2:19:24 PM. (outdated)

```

2. Wenn Sie FLINK JOB wählen, gelangen Sie zur Flink Web Console in einer anderen Registerkarte Ihres Browsers.



The screenshot shows the Apache Flink Dashboard interface. On the left is a navigation sidebar with options: Overview, Jobs (Running Jobs, Completed Jobs), Task Managers, Job Manager, and Submit New Job. The main content area displays:

- Available Task Slots:** 0. Total Task Slots: 0, Task Managers: 0.
- Running Jobs:** 0. Finished: 2, Canceled: 0, Failed: 0.
- Running Job List:** A table with columns: Job Name, Start Time, Duration, End Time, Tasks, Status. It shows "No Data".
- Completed Job List:** A table with columns: Job Name, Start Time, Duration, End Time, Tasks, Status.
 

Job Name	Start Time	Duration	End Time	Tasks	Status
Flink Streaming Job	2023-05-04 14:20:56	8s	2023-05-04 14:21:04	2 / 2	FINISHED
Flink Java Job at Thu May 04 08:49:10 UTC 2023	2023-05-04 14:19:11	12s	2023-05-04 14:19:23	3 / 3	FINISHED

## Versionsverlauf von Flink

In der folgenden Tabelle sind die Version von Flink aufgeführt, die in jeder Release-Version von Amazon EMR enthalten ist, zusammen mit den Komponenten, die mit der Anwendung installiert wurden. Informationen zu den Komponentenversionen in den einzelnen Versionen finden Sie im Abschnitt Komponentenversion für Ihre Version in [Amazon-EMR-7.x-Versionen](#), [Amazon-EMR-6.x-Versionen](#) oder [Amazon-EMR-5.x-Versionen](#).

### Die Flink-Versionsinformationen

Amazon-EMR-Version sbezeichnung	Flink-Version	Mit Flink installierte Komponenten
emr-5.36,2	1.14.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya

Amazon-EMR-Version sbezeichnung	Flink-Version	Mit Flink installierte Komponenten
		rn-timeline-server, flink-client, flink-jobmanager-config
emr-7.1.0	1.18.1	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config, hudi, delta
emr-7.0.0	1.18.0	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config, hudi, delta-standalone-connectors

Amazon-EMR-Version sbezeichnung	Flink-Version	Mit Flink installierte Komponenten
emr-6.15.0	1.17.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config, hudi, delta-standalone-connectors
emr-6.14.0	1.17.1	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config, hudi, delta-standalone-connectors

Amazon-EMR-Version sbezeichnung	Flink-Version	Mit Flink installierte Komponenten
emr-6.13.0	1.17.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config, hudi, delta-standalone-connectors
emr-6.12.0	1.17.0	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config, hudi, delta-standalone-connectors

Amazon-EMR-Version sbezeichnung	Flink-Version	Mit Flink installierte Komponenten
emr-6.11.1	1.16.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config, hudi, delta-standalone-connectors
emr-6.11.0	1.16.0	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config, hudi, delta-standalone-connectors
emr-6.10.1	1.16.0	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config, hudi

Amazon-EMR-Version sbezeichnung	Flink-Version	Mit Flink installierte Komponenten
emr-6.10.0	1.16.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config, hudi
emr-6.9.1	1.15.2	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config, hudi
emr-6.9.0	1.15.2	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config, hudi

Amazon-EMR-Version sbezeichnung	Flink-Version	Mit Flink installierte Komponenten
emr-6.8.1	1.15.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config, hudi
emr-6.8.0	1.15.1	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config, hudi
emr-6.7.0	1.14.2	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config, hudi

Amazon-EMR-Version sbezeichnung	Flink-Version	Mit Flink installierte Komponenten
emr-5.36.1	1.14.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config
emr-5.36.0	1.14.2	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config
emr-6.6.0	1.14.2	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config, hudi



Amazon-EMR-Version sbezeichnung	Flink-Version	Mit Flink installierte Komponenten
emr-5.35.0	1.14.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config
emr-6.5.0	1.14.0	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config, hudi
emr-6.4.0	1.13.1	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config, hudi

Amazon-EMR-Version sbezeichnung	Flink-Version	Mit Flink installierte Komponenten
emr-6.3.1	1.12.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config
emr-6.3.0	1.12.1	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config
emr-6.2.1	1.11.2	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config

Amazon-EMR-Version sbezeichnung	Flink-Version	Mit Flink installierte Komponenten
emr-6.2.0	1.11.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config
emr-6.1.1	1.11.0	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-6.1.0	1.11.0	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client

Amazon-EMR-Version sbezeichnung	Flink-Version	Mit Flink installierte Komponenten
emr-5.34.0	1.13.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config
emr-5.33.1	1.12.1	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config
emr-5.33.0	1.12.1	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config

Amazon-EMR-Version sbezeichnung	Flink-Version	Mit Flink installierte Komponenten
emr-5.32.1	1.11.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config
emr-5.32.0	1.11.2	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config
emr-5.31.1	1.11.0	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config

Amazon-EMR-Version sbezeichnung	Flink-Version	Mit Flink installierte Komponenten
emr-5.31.0	1.11.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config
emr-5.30.2	1.10.0	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.30.1	1.10.0	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client

Amazon-EMR-Version sbezeichnung	Flink-Version	Mit Flink installierte Komponenten
emr-5.30.0	1.10.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.29.0	1.9.1	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.28.1	1.9.0	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client

Amazon-EMR-Version sbezeichnung	Flink-Version	Mit Flink installierte Komponenten
emr-5.28.0	1.9.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.27.1	1.8.1	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.27.0	1.8.1	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client



Amazon-EMR-Version sbezeichnung	Flink-Version	Mit Flink installierte Komponenten
emr-5.26.0	1.8.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.25.0	1.8.0	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.24.1	1.8.0	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client

Amazon-EMR-Version sbezeichnung	Flink-Version	Mit Flink installierte Komponenten
emr-5.24.0	1.8.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.23.1	1.7.1	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.23.0	1.7.1	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client

Amazon-EMR-Version sbezeichnung	Flink-Version	Mit Flink installierte Komponenten
emr-5.22.0	1.7.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.21.2	1.7.0	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.21.1	1.7.0	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client

Amazon-EMR-Version sbezeichnung	Flink-Version	Mit Flink installierte Komponenten
emr-5.21.0	1.7.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.20.1	1.6.2	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.20.0	1.6.2	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client

Amazon-EMR-Version sbezeichnung	Flink-Version	Mit Flink installierte Komponenten
emr-5.19.1	1.6.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.19.0	1.6.1	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.18.1	1.6.0	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client

Amazon-EMR-Version sbezeichnung	Flink-Version	Mit Flink installierte Komponenten
emr-5.18.0	1.6.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.17.2	1.5.2	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.17.1	1.5.2	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client

Amazon-EMR-Version sbezeichnung	Flink-Version	Mit Flink installierte Komponenten
emr-5.17.0	1.5.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.16.1	1.5.0	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.16.0	1.5.0	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client

Amazon-EMR-Version sbezeichnung	Flink-Version	Mit Flink installierte Komponenten
emr-5.15.1	1.4.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.15.0	1.4.2	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.14.2	1.4.2	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client



Amazon-EMR-Version sbezeichnung	Flink-Version	Mit Flink installierte Komponenten
emr-5.14.1	1.4.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.14.0	1.4.2	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.13.1	1.4.0	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client

Amazon-EMR-Version sbezeichnung	Flink-Version	Mit Flink installierte Komponenten
emr-5.13.0	1.4.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.12.3	1.4.0	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.12.2	1.4.0	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client

Amazon-EMR-Version sbezeichnung	Flink-Version	Mit Flink installierte Komponenten
emr-5.12.1	1.4.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.12.0	1.4.0	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.11.4	1.3.2	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client

Amazon-EMR-Version sbezeichnung	Flink-Version	Mit Flink installierte Komponenten
emr-5.11.3	1.3.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.11.2	1.3.2	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.11.1	1.3.2	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client

Amazon-EMR-Version sbezeichnung	Flink-Version	Mit Flink installierte Komponenten
emr-5.11.0	1.3.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.10.1	1.3.2	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.10.0	1.3.2	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client

Amazon-EMR-Version sbezeichnung	Flink-Version	Mit Flink installierte Komponenten
emr-5.9.1	1.3.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.9.0	1.3.2	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.8.3	1.3.1	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client

Amazon-EMR-Version sbezeichnung	Flink-Version	Mit Flink installierte Komponenten
emr-5.8.2	1.3.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.8.1	1.3.1	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.8.0	1.3.1	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client

Amazon-EMR-Version sbezeichnung	Flink-Version	Mit Flink installierte Komponenten
emr-5.7.1	1.3.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.7.0	1.3.0	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.6.1	1.2.1	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client



Amazon-EMR-Version sbezeichnung	Flink-Version	Mit Flink installierte Komponenten
emr-5.6.0	1.2.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.5.4	1.2.0	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, flink-client
emr-5.5.3	1.2.0	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, flink-client

Amazon-EMR-Version sbezeichnung	Flink-Version	Mit Flink installierte Komponenten
emr-5.5.2	1.2.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, flink-client
emr-5.5.1	1.2.0	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, flink-client
emr-5.5.0	1.2.0	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, flink-client

Amazon-EMR-Version sbezeichnung	Flink-Version	Mit Flink installierte Komponenten
emr-5.4.1	1.2.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, flink-client
emr-5.4.0	1.2.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, flink-client
emr-5.3.2	1.1.4	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, flink-client

Amazon-EMR-Version sbezeichnung	Flink-Version	Mit Flink installierte Komponenten
emr-5.3.1	1.1.4	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, flink-client
emr-5.3.0	1.1.4	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, flink-client
emr-5.2.3	1.1.3	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, flink-client

Amazon-EMR-Version sbezeichnung	Flink-Version	Mit Flink installierte Komponenten
emr-5.2.2	1.1.3	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, flink-client
emr-5.2.1	1.1.3	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, flink-client
emr-5.2.0	1.1.3	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, flink-client

Amazon-EMR-Version sbezeichnung	Flink-Version	Mit Flink installierte Komponenten
emr-5.1.1	1.1.3	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, flink-client
emr-5.1.0	1.1.3	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, flink-client

# Ganglia

## Note

Die letzte Version von Amazon EMR, die Ganglia enthielt, war Amazon EMR 6.15.0. Um Ihren Cluster zu überwachen, enthalten Versionen über 6.15.0 den [CloudWatch Amazon-Agenten](#).

Das Open-Source-Projekt Ganglia ist ein skalierbares, verteiltes System zur Überwachung von Clustern und Grids, das zugleich die Auswirkungen auf die Leistung minimiert. Wenn Sie Ganglia in Ihrem Cluster aktivieren, können Sie Berichte erstellen und die Leistung des Clusters als Ganzes betrachten. Ebenso können Sie die Leistung einzelner Knoten-Instances überprüfen. Ganglia ist außerdem zur Aufnahme und Visualisierung von Hadoop- und Spark-Metriken konfiguriert. Weitere Informationen zum Open-Source-Projekt Ganglia finden Sie unter <http://ganglia.info/>.

Wenn Sie die Ganglia-Web-Benutzeroberfläche in einem Browser anzeigen, sehen Sie eine Übersicht über die Cluster-Leistung mit Diagrammen, die Details zur Last, Arbeitsspeicherauslastung, CPU-Auslastung und zum Netzwerkverkehr des Clusters enthalten. Unter den Cluster-Statistiken werden Diagramme für jeden einzelnen Server im Cluster angegeben.

Die folgende Tabelle listet die Version von Ganglia auf, die in der neuesten Version der Amazon-EMR-6.x-Serie enthalten ist, zusammen mit den Komponenten, die Amazon EMR mit Ganglia installiert.

Die Version der Komponenten, die mit Ganglia in dieser Version installiert wurden, finden Sie unter [Komponentenversionen der Version 6.15.0](#).

Ganglia-Versionsinformationen für emr-6.15.0

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
emr-6.15.0	Ganglia 3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
		server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Die folgende Tabelle listet die Version von Ganglia auf, die in der neuesten Version der Amazon-EMR-5.x-Serie enthalten ist, zusammen mit den Komponenten, die Amazon EMR mit Ganglia installiert.

Informationen zur Version der Komponenten, die in dieser Version mit Ganglia installiert wurden, finden Sie unter [Komponentenversionen von Version 5.36.2](#).

#### Ganglia-Versionsinformationen für emr-5.36.2

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
emr-5.36.2	Ganglia 3.7.2	emrfs, emr-goodies, gangliamonitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

#### Themen

- [Erstellen eines Clusters mit Ganglia](#)
- [Anzeigen von Ganglia-Metriken](#)
- [Hadoop- und Spark-Metriken in Ganglia](#)
- [Versionsverlauf von Ganglia](#)



# Erstellen eines Clusters mit Ganglia

## Note

Die letzte Version von Amazon EMR, die Ganglia enthielt, war Amazon EMR 6.15.0. Um Ihren Cluster zu überwachen, enthalten Versionen über 6.15.0 den [CloudWatch Amazon-Agenten](#).

So erstellen Sie einen Cluster mit Ganglia über die Konsole

1. Navigieren Sie zur neuen Amazon-EMR-Konsole und wählen Sie in der Seitennavigation die Option Zur alten Konsole wechseln aus. Weitere Informationen darüber, was Sie erwartet, wenn Sie zur alten Konsole wechseln, finden Sie unter [Verwenden der alten Konsole](#).
2. Wählen Sie Cluster erstellen.
3. Wählen Sie unter Software configuration (Softwarekonfiguration) entweder All Applications (Alle Anwendungen), Core Hadoop oder Spark aus.
4. Fahren Sie mit der Erstellung des Clusters mit Konfigurationen nach Bedarf fort.

Um Ganglia zu einem Cluster hinzuzufügen, verwenden Sie AWS CLI

In der AWS CLI können Sie Ganglien zu einem Cluster hinzufügen, indem Sie `create-cluster` den `--applications` Parameter verwenden. Wenn Sie nur Ganglia mit dem Parameter `--applications` angeben, ist Ganglia die einzige installierte Anwendung.

- Geben Sie den folgenden Befehl ein, um Ganglia hinzuzufügen, wenn Sie einen Cluster erstellen. Ersetzen Sie *myKey* durch den Namen Ihres EC2-Schlüsselpaars.

## Note

Linux-Zeilenfortsetzungszeichen (`\`) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (`^`).

```
aws emr create-cluster --name "Spark cluster with Ganglia" --release-label  
emr-6.15.0; \
```

```
--applications Name=Spark Name=Ganglia \  
--ec2-attributes KeyName=myKey --instance-type m5.xlarge \  
--instance-count 3 --use-default-roles
```

Wenn Sie die Instance-Anzahl ohne den Parameter `--instance-groups` angeben, wird ein einzelner Master-Knoten gestartet. Die verbleibenden Instances werden als Core-Knoten aufgerufen. Alle Knoten verwenden den im Befehl angegebenen Instance-Typ.

#### Note

Wenn Sie zuvor nicht die standardmäßige EMR-Servicerolle und das EC2-Instance-Profil erstellt haben, geben Sie `aws emr create-default-roles` ein, um sie zu erstellen, bevor Sie den Unterbefehl `create-cluster` eingeben.

Weitere Informationen zur Verwendung von Amazon EMR-Befehlen finden Sie AWS CLI unter <https://docs.aws.amazon.com/cli/latest/reference/emr>.

## Anzeigen von Ganglia-Metriken

#### Note

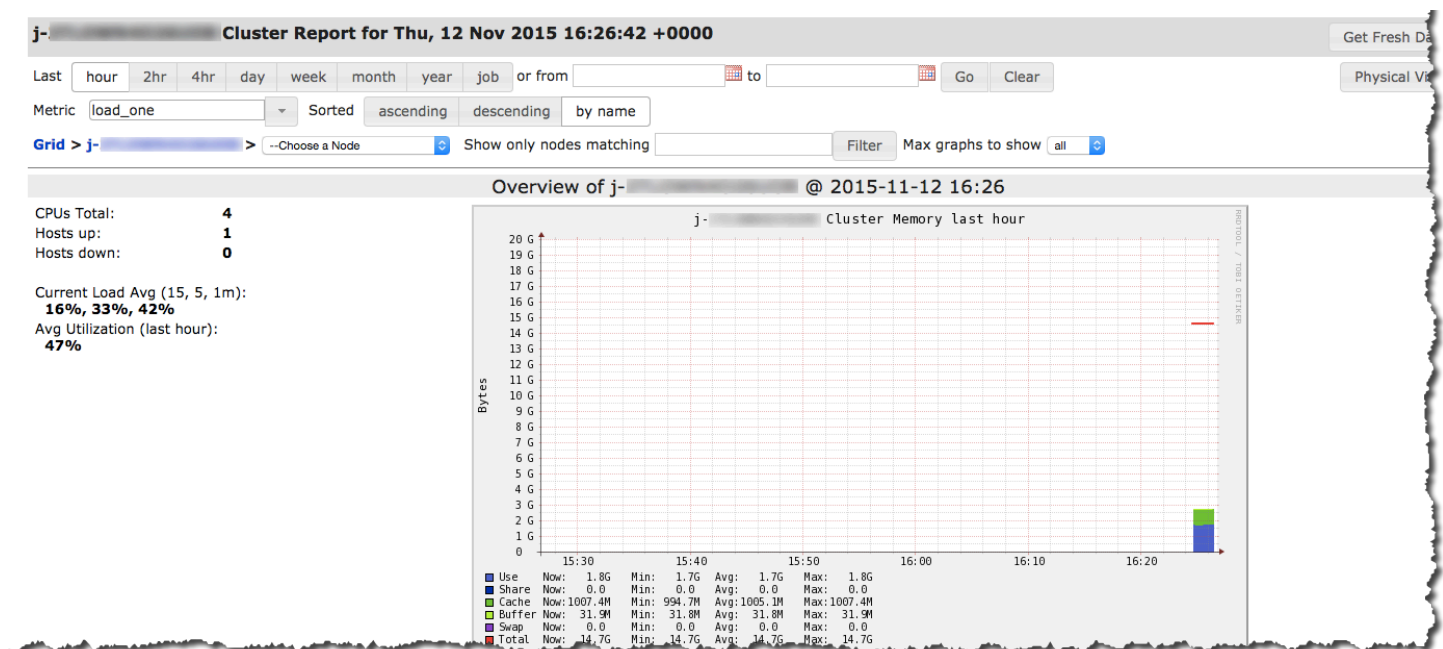
Die letzte Version von Amazon EMR, die Ganglia enthielt, war Amazon EMR 6.15.0. Um Ihren Cluster zu überwachen, enthalten Versionen über 6.15.0 den [CloudWatch Amazon-Agenten](#).

Ganglia bietet eine webbasierte Benutzeroberfläche, die Sie verwenden können, um Metriken Ganglia anzuzeigen. Wenn Sie Ganglia auf Amazon EMR ausführen, wird die Web-Schnittstelle auf dem Hauptknoten ausgeführt und kann über die Port-Weiterleitung (Erstellung eines SSH-Tunnels) angezeigt werden. Weitere Informationen zum Anzeigen von Webbenutzeroberflächen auf finden Sie unter [Auf Amazon-EMR-Clustern gehostete Webschnittstellen anzeigen](#) im Verwaltungshandbuch für Amazon EMR.

So zeigen Sie die Ganglia Weboberfläche an

1. Verwenden Sie SSH, um einen Tunnel zum Master-Knoten und eine sichere Verbindung zu erstellen. Weitere Informationen zum Erstellen eines SSH-Tunnels zum Hauptknoten finden

- Sie unter [Option 2, Teil 1: Einrichten eines SSH-Tunnels zum Hauptknoten mit dynamischer Portweiterleitung](#) im Verwaltungshandbuch für Amazon EMR.
2. Installieren Sie einen Webbrowser mit einem Proxy-Tool, z. B. dem FoxyProxy Plug-in für Firefox, um einen SOCKS-Proxy für Domains des Typs `*ec2*.amazonaws.com*` zu erstellen. Weitere Informationen finden Sie unter [Option 2, Teil 2: Proxy-Einstellungen konfigurieren, um auf dem Hauptknoten gehostete Websites anzeigen zu lassen](#) im Verwaltungshandbuch für Amazon EMR.
  3. Wenn der Proxy gesetzt und die SSH-Verbindung geöffnet ist, können Sie die Ganglia-Benutzeroberfläche anzeigen, indem Sie ein Browserfenster mit `http://master-public-dns-name/ganglia/` öffnen, in dem sich die öffentliche DNS-Adresse des Masterservers im EMR-Cluster `master-public-dns-name` befindet.



## Hadoop- und Spark-Metriken in Ganglia

### Note

Die letzte Version von Amazon EMR, die Ganglia enthielt, war Amazon EMR 6.15.0. Um Ihren Cluster zu überwachen, enthalten Versionen über 6.15.0 den [CloudWatch Amazon-Agenten](#).

Ganglia meldet Hadoop-Metriken für jede Instance. Die verschiedenen Arten von Metriken sind nach Kategorien mit einem Präfix versehen: verteiltes Dateisystem (dfs.\*), Java Virtual Machine (jvm.\*), (mapred.\*) und Remote Procedure Calls MapReduce (rpc.\*).

Auf YARN basierende Ganglia-Metriken wie Spark und Hadoop sind für die EMR-Versionen 4.4.0 und 4.5.0 nicht verfügbar. Verwenden Sie eine höhere Version, um diese Metriken zu verwenden.

Ganglia-Metriken für Spark haben in der Regel Präfixe für YARN-Anwendungs-ID und Spark-DAGScheduler. Die Präfixe haben also folgende Form:

- DAGScheduler.\*
- application\_XXXXXXXXXX\_XXXX.driver.\*
- application\_XXXXXXXXXX\_XXXX.executor.\*

## Versionsverlauf von Ganglia

### Note

Die letzte Version von Amazon EMR, die Ganglia enthielt, war Amazon EMR 6.15.0. Um Ihren Cluster zu überwachen, enthalten Versionen über 6.15.0 den [CloudWatch Amazon-Agenten](#).

In der folgenden Tabelle sind die Version von Ganglia aufgeführt, die in jeder Release-Version von Amazon EMR enthalten ist, zusammen mit den Komponenten, die mit der Anwendung installiert wurden. Informationen zu den Komponentenversionen in den einzelnen Versionen finden Sie im Abschnitt [Komponentenversion für Ihre Version in Amazon-EMR-7.x-Versionen](#), [Amazon-EMR-6.x-Versionen](#) oder [Amazon-EMR-5.x-Versionen](#).

### Informationen zur Ganglia-Version

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
emr-5.36.2	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode,

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
		hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver
emr-6.15.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver
emr-6.14.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
emr-6.13.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-6.12.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-6.11.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
emr-6.11.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-6.10.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-6.10.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
emr-6.9.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-6.9.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-6.8.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver



Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
emr-6.8.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-6.7.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.36.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
emr-5.36.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-6.6.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.35.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
emr-6.5.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-6.4.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-6.3.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
emr-6.3.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-6.2.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-6.2.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
emr-6.1.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-6.1.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-6.0.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
emr-6.0.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.34.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.33.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
emr-5.33.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.32.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.32.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
emr-5.31.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.31.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.30.2	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver



Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
emr-5.30.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.30.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.29.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
emr-5.28.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.28.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.27.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
emr-5.27.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.26.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.25.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
emr-5.24.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.24.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.23.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
emr-5.23.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.22.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.21.2	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
emr-5.21.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.21.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.20.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
emr-5.20.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.19.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.19.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
emr-5.18.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.18.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.17.2	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver



Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
emr-5.17.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.17.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.16.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
emr-5.16.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.15.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.15.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
emr-5.14.2	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.14.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.14.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
emr-5.13.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.13.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.12.3	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
emr-5.12.2	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.12.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.12.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
emr-5.11.4	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.11.3	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.11.2	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
emr-5.11.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.11.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.10.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
emr-5.10.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.9.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.9.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver



Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
emr-5.8.3	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.8.2	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.8.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
emr-5.8.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.7.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.7.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
emr-5.6.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.6.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.5.4	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
emr-5.5.3	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver
emr-5.5.2	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver
emr-5.5.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
emr-5.5.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-5.4.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-5.4.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
emr-5.3.2	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver
emr-5.3.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver
emr-5.3.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
emr-5.2.3	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-5.2.2	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-5.2.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
emr-5.2.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver
emr-5.1.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver
emr-5.1.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver



Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
emr-5.0.3	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-5.0.2	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-5.0.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
emr-5.0.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver
emr-4.9.6	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver
emr-4.9.5	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
emr-4.9.4	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-4.9.3	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-4.9.2	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
emr-4.9.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver
emr-4.8.5	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver
emr-4.8.4	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
emr-4.8.3	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-4.8.2	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-4.8.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
emr-4.8.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver
emr-4.7.4	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver
emr-4.7.3	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
emr-4.7.2	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-4.7.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-4.7.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver

Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
emr-4.6.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-4.6.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-4.5.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver



Amazon-EMR-Version sbezeichnung	Ganglia-Version	Mit Ganglia installierte Komponenten
emr-4.4.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-4.3.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-4.2.0	3.6.0	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver

# Apache Hadoop

[Apache Hadoop](#) ist ein Open-Source-Java-Software-Framework, das die Verarbeitung von großen Datenmengen mithilfe von Instance-Clustern unterstützt. Hadoop kann auf einer einzigen Instance oder auch auf Tausenden von Instances ausgeführt werden. Hadoop verwendet verschiedene Verarbeitungsmodelle wie MapReduce und Tez, um die Verarbeitung auf mehrere Instanzen zu verteilen, und verwendet außerdem ein verteiltes Dateisystem namens HDFS, um Daten auf mehreren Instanzen zu speichern. Hadoop überwacht den Zustand der Instances im Cluster und kann nach dem Ausfall eines oder mehrerer Knoten eine Wiederherstellung durchführen. Auf diese Weise bietet Hadoop nicht nur erhöhte Datenverarbeitungs- und Speicherkapazitäten, sondern auch eine hohe Verfügbarkeit. Weitere Informationen finden Sie in der [Hadoop-Dokumentation](#).

Die folgende Tabelle listet die Version von Hadoop auf, die in der neuesten Version der Amazon-EMR-7.x-Serie enthalten ist, zusammen mit den Komponenten, die Amazon EMR mit Tez installiert.

[Informationen zur Version der Komponenten, die in dieser Version mit Hadoop installiert wurden, finden Sie unter Komponentenversionen der Version 7.1.0.](#)

## Hadoop-Versionsinformationen für emr-7.1.0

Amazon-EMR-Version sbezeichnung	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-7.1.0	Hadoop 3.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Die folgende Tabelle listet die Version von Hadoop auf, die in der neuesten Version der Amazon-EMR-6.x-Serie enthalten ist, zusammen mit den Komponenten, die Amazon EMR mit Hadoop installiert.

Die Version der Komponenten, die mit Hadoop in dieser Version installiert wurden, finden Sie unter [Komponentenversionen der Version 6.15.0](#).

#### Hadoop-Versionsinformationen für emr-6.15.0

Amazon-EMR-Version sbezeichnung	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-6.15.0	Hadoop 3.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Die folgende Tabelle listet die Version von Hadoop auf, die in der neuesten Version der Amazon-EMR-5.x-Serie enthalten ist, zusammen mit den Komponenten, die Amazon EMR mit Tez installiert.

[Informationen zur Version der Komponenten, die in dieser Version mit Hadoop installiert wurden, finden Sie unter Komponentenversionen von Version 5.36.2.](#)

#### Hadoop-Versionsinformationen für emr-5.36.2

Amazon-EMR-Version sbezeichnung	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.36.2	Hadoop 2.10.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-

Amazon-EMR-Version sbezeichnung	Hadoop-Version	Mit Hadoop installierte Komponenten
		library, hadoop-hdfs-nameno de, hadoop-httfs-server, hadoop-kms-server, hadoop- mapred, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server

Ab Amazon EMR 5.18.0 können Sie das Amazon-EMR-Artefakt-Repository verwenden, um Ihren Auftragscode anhand der genauen Versionen von Bibliotheken und Abhängigkeiten zu erstellen, die mit bestimmten Amazon-EMR-Versionen verfügbar sind. Weitere Informationen finden Sie unter [Überprüfen von Abhängigkeiten mithilfe des Amazon-EMR-Artefakt-Repositorys](#).

#### Themen

- [Konfigurieren von Hadoop](#)
- [Transparente Verschlüsselung in HDFS auf Amazon EMR](#)
- [Erstellen oder Ausführen einer Hadoop-Anwendung](#)
- [Aktivieren Sie die Erkennung ungleichmäßiger Speicherzugriffe für YARN-Container](#)
- [Hadoop-Versionsverlauf](#)

## Konfigurieren von Hadoop

In den folgenden Abschnitten werden die Standardkonfigurationseinstellungen für Hadoop-Daemons, Aufgaben und HDFS aufgeführt.

#### Themen

- [Aufgabenkonfiguration](#)
- [Hadoop-Daemon-Konfigurationseinstellungen](#)
- [HDFS-Konfiguration](#)

## Aufgabenkonfiguration

Sie können Konfigurationsvariablen festlegen, um die Leistung Ihrer MapReduce Jobs zu optimieren. Dieser Abschnitt enthält die Standardwerte für wichtige Einstellungen. Standardwerte variieren je nach dem EC2-Instance-Typ des im Cluster verwendeten Knoten. HBase steht bei Verwendung von Amazon-EMR-Version 4.6.0 und höher zur Verfügung. Unterschiedliche Standardeinstellungen werden verwendet, wenn HBase installiert ist. Diese Werte werden zusammen mit den anfänglichen Standardwerten bereitgestellt.

Hadoop 2 verwendet die zwei Parameter `mapreduce.map.java.opts` und `mapreduce.reduce.java.opts` zum Konfigurieren des Arbeitsspeichers für die Karte und zum entsprechenden Verringern der JVMs. Diese ersetzen die einzelne `mapreduce.map.java.opts`-Konfigurationsoption von früheren Hadoop-Versionen.

Gleichermaßen ersetzt `mapred.job.jvm.num.tasks` `mapred.job.reuse.jvm.num.tasks` in Hadoop 2.7.2 und höher. Amazon EMR setzt diesen Wert unabhängig vom EC2-Instance-Typ auf 20. Sie können diese Einstellung mithilfe der `mapred-site`-Konfigurationsklassifizierung überschreiben. Durch Festlegen eines Werts auf `-1` wird angegeben, dass eine JVM für eine unendliche Anzahl an Aufgaben innerhalb eines einzelnen Auftrags wiederverwendet werden kann. Der Wert `1` gibt an, dass eine neue JVM für jede Aufgabe erzeugt wird.

Beispiel: Um den Wert von `mapred.job.jvm.num.tasks` auf `-1` festzulegen, können Sie eine Datei mit dem folgenden Inhalt erstellen:

```
[
  {
    "Classification": "mapred-site",
    "Properties": {
      "mapred.job.jvm.num.tasks": "-1"
    }
  }
]
```

Wenn Sie den `create-cluster` Befehl oder den `modify-instance-groups` Befehl aus dem verwenden AWS CLI, können Sie dann auf die JSON-Konfigurationsdatei verweisen. Im folgenden Beispiel wird die Konfigurationsdatei unter `myConfig.json` gespeichert und in Amazon S3 aufbewahrt.

**Note**

Linux-Zeilenumbruchzeichen (\) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (^).

```
aws emr create-cluster --release-label emr-7.1.0 --instance-type m5.xlarge \  
--instance-count 3 --applications Name=Hadoop --configurations https://  
s3.amazonaws.com/mybucket/myfolder/myConfig.json \  
--use-default-roles
```

Sie können die unten aufgeführten Standardwerte mithilfe der `mapred-site-`Konfigurationsklassifizierung auf dieselbe Weise ändern und mehrere Werte und mehrere Konfigurationsklassifizierungen mit einer einzelnen JSON-Datei festlegen. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Ab Amazon-EMR-Version 5.21.0 können Sie Cluster-Konfigurationen überschreiben und zusätzliche Konfigurationsklassifikationen für jede Instance-Gruppe in einem ausgeführten Cluster angeben. Dazu verwenden Sie die Amazon EMR-Konsole, das AWS Command Line Interface (AWS CLI) oder das AWS SDK. Weitere Informationen finden Sie unter [Angaben einer Konfiguration für eine Instance-Gruppe in einem aktiven Cluster](#).

## Standardwerte für die Aufgabenkonfigurationseinstellungen

### Instance-Typen

- [C1-Instances](#)
- [C3-Instances](#)
- [C4-Instances](#)
- [C5-Instances](#)
- [C5a-Instances](#)
- [C5ad-Instances](#)
- [C5d-Instances](#)
- [C5n-Instances](#)
- [C6a-Instances](#)

- [C6g-Instances](#)
- [C6gd-Instances](#)
- [C6gn-Instances](#)
- [C6i-Instances](#)
- [C6id-Instances](#)
- [C6in-Instances](#)
- [c7a-Instances](#)
- [C7g-Instances](#)
- [C7gd-Instances](#)
- [C7gn-Instances](#)
- [C7i-Instances](#)
- [D2-Instances](#)
- [D3-Instances](#)
- [D3en-Instances](#)
- [G3-Instances](#)
- [G3s-Instances](#)
- [G4dn-Instances](#)
- [G5-Instances](#)
- [H1-Instances](#)
- [I2-Instances](#)
- [I3-Instances](#)
- [I3en-Instances](#)
- [I4g-Instances](#)
- [I4i-Instances](#)
- [Im4gn-Instances](#)
- [Is4gen-Instances](#)
- [M1-Instances](#)
- [M2-Instances](#)
- [M3-Instances](#)
- [M4-Instances](#)

- [M5-Instances](#)
- [M5a-Instances](#)
- [M5ad-Instances](#)
- [M5d-Instances](#)
- [M5dn-Instances](#)
- [M5n-Instances](#)
- [M5zn-Instances](#)
- [M6a-Instances](#)
- [M6g-Instances](#)
- [M6gd-Instances](#)
- [M6i-Instances](#)
- [M6id-Instances](#)
- [M6idn-Instances](#)
- [M6in-Instances](#)
- [M7a-Instances](#)
- [M7g-Instances](#)
- [M7gd-Instances](#)
- [M7i-Instances](#)
- [M7i-flex-Instances](#)
- [P2-Instances](#)
- [P3-Instances](#)
- [P5-Instances](#)
- [R3-Instances](#)
- [R4-Instances](#)
- [R5-Instances](#)
- [R5a-Instances](#)
- [R5ad-Instances](#)
- [r5b-Instances](#)
- [R5d-Instances](#)
- [R5dn-Instances](#)



- [R5n-Instances](#)
- [R6a-Instances](#)
- [R6g-Instances](#)
- [R6gd-Instances](#)
- [R6i-Instances](#)
- [R6id-Instances](#)
- [R6idn-Instances](#)
- [R6in-Instances](#)
- [R7a-Instances](#)
- [R7g-Instances](#)
- [R7gd-Instances](#)
- [r7i-Instances](#)
- [R7iz-Instances](#)
- [X1-Instances](#)
- [X1e-Instances](#)
- [X2gd-Instances](#)
- [x2idn-Instances](#)
- [X2iedn-Instances](#)
- [z1d-Instances](#)

## C1-Instances

### c1.medium

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx288m</code>	<code>-Xmx288m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx288m</code>	<code>-Xmx288m</code>
<code>mapreduce.map.memory.mb</code>	512	512
<code>mapreduce.reduce.memory.mb</code>	512	512

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.app.mapreduce.am.resource.mb</code>	512	512
<code>yarn.scheduler.minimum-allocation-mb</code>	256	256
<code>yarn.scheduler.maximum-allocation-mb</code>	512	512
<code>yarn.nodemanager.resource.memory-mb</code>	1024	512

## c1.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx864m</code>	<code>-Xmx864m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx1536m</code>	<code>-Xmx1536m</code>
<code>mapreduce.map.memory.mb</code>	1024	1024
<code>mapreduce.reduce.memory.mb</code>	2048	2048
<code>yarn.app.mapreduce.am.resource.mb</code>	2048	2048
<code>yarn.scheduler.minimum-allocation-mb</code>	256	256
<code>yarn.scheduler.maximum-allocation-mb</code>	2048	2560
<code>yarn.nodemanager.resource.memory-mb</code>	5120	2560

## C3-Instances

## c3.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1126m	-Xmx1126m
mapreduce.java.opts	-Xmx2252m	-Xmx2252m
mapreduce.map.memory.mb	1 408	1 408
mapreduce.reduce.memory.mb	2816	2816
yarn.app.mapreduce.am.resource.mb	2816	2816
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	5632	2816
yarn.nodemanager.resource.memory-mb	5632	2816

## c3.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1152m	-Xmx1152m
mapreduce.java.opts	-Xmx2304m	-Xmx2304m
mapreduce.map.memory.mb	1440	1440
mapreduce.reduce.memory.mb	2880	2880

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.app.mapreduce.am.resource.mb</code>	2880	2880
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	11520	5760
<code>yarn.nodemanager.resource.memory-mb</code>	11520	5760

## c3.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx1152m</code>	<code>-Xmx1152m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2304m</code>	<code>-Xmx2304m</code>
<code>mapreduce.map.memory.mb</code>	1440	1440
<code>mapreduce.reduce.memory.mb</code>	2880	2880
<code>yarn.app.mapreduce.am.resource.mb</code>	2880	2880
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23040	11520
<code>yarn.nodemanager.resource.memory-mb</code>	23040	11520

## c3.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx1331m</code>	<code>-Xmx1331m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2662m</code>	<code>-Xmx2662m</code>
<code>mapreduce.map.memory.mb</code>	1664	1664
<code>mapreduce.reduce.memory.mb</code>	3328	3328
<code>yarn.app.mapreduce.am.resource.mb</code>	3328	3328
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	53248	26624
<code>yarn.nodemanager.resource.memory-mb</code>	53248	26624

## C4-Instances

## c4.large

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx717m</code>	<code>-Xmx717m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx1434m</code>	<code>-Xmx1434m</code>
<code>mapreduce.map.memory.mb</code>	896	896
<code>mapreduce.reduce.memory.mb</code>	1792	1792

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.app.mapreduce.am.resource.mb	1792	1792
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	1792	896
yarn.nodemanager.resource.memory-mb	1792	896

## c4.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1126m	-Xmx1126m
mapreduce.java.opts	-Xmx2252m	-Xmx2252m
mapreduce.map.memory.mb	1 408	1 408
mapreduce.reduce.memory.mb	2816	2816
yarn.app.mapreduce.am.resource.mb	2816	2816
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	5632	2816
yarn.nodemanager.resource.memory-mb	5632	2816

## c4.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx1152m</code>	<code>-Xmx1152m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2304m</code>	<code>-Xmx2304m</code>
<code>mapreduce.map.memory.mb</code>	1440	1440
<code>mapreduce.reduce.memory.mb</code>	2880	2880
<code>yarn.app.mapreduce.am.resource.mb</code>	2880	2880
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	11520	5760
<code>yarn.nodemanager.resource.memory-mb</code>	11520	5760

## c4.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx1152m</code>	<code>-Xmx1152m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2304m</code>	<code>-Xmx2304m</code>
<code>mapreduce.map.memory.mb</code>	1440	1440
<code>mapreduce.reduce.memory.mb</code>	2880	2880
<code>yarn.app.mapreduce.am.resource.mb</code>	2880	2880

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23040	11520
<code>yarn.nodemanager.resource.memory-mb</code>	23040	11520

## c4.8xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx1183m</code>	<code>-Xmx1183m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2366m</code>	<code>-Xmx2366m</code>
<code>mapreduce.map.memory.mb</code>	1479	1479
<code>mapreduce.reduce.memory.mb</code>	2958	2958
<code>yarn.app.mapreduce.am.resource.mb</code>	2958	2958
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	53248	26624
<code>yarn.nodemanager.resource.memory-mb</code>	53248	26624



## C5-Instances

## c5.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1229m	-Xmx1229m
mapreduce.java.opts	-Xmx2458m	-Xmx2458m
mapreduce.map.memory.mb	1536	1536
mapreduce.reduce.memory.mb	3072	3072
yarn.app.mapreduce.am.resource.mb	3072	3072
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	6144	3072
yarn.nodemanager.resource.memory-mb	6144	3072

## c5.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1229m	-Xmx1229m
mapreduce.java.opts	-Xmx2458m	-Xmx2458m
mapreduce.map.memory.mb	1536	1536
mapreduce.reduce.memory.mb	3072	3072

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.app.mapreduce.am.resource.mb	3072	3072
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	12288	6144
yarn.nodemanager.resource.memory-mb	12288	6144

## c5.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1229m	-Xmx1229m
mapreduce.java.opts	-Xmx2458m	-Xmx2458m
mapreduce.map.memory.mb	1536	1536
mapreduce.reduce.memory.mb	3072	3072
yarn.app.mapreduce.am.resource.mb	3072	3072
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	24576	12288
yarn.nodemanager.resource.memory-mb	24576	12288

## c5.9xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx1456m</code>	<code>-Xmx1456m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2912m</code>	<code>-Xmx2912m</code>
<code>mapreduce.map.memory.mb</code>	1820	1820
<code>mapreduce.reduce.memory.mb</code>	3640	3640
<code>yarn.app.mapreduce.am.resource.mb</code>	3640	3640
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	65536	32768
<code>yarn.nodemanager.resource.memory-mb</code>	65536	32768

## c5.12xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx1502m</code>	<code>-Xmx1502m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx3004m</code>	<code>-Xmx3004m</code>
<code>mapreduce.map.memory.mb</code>	1877	1877
<code>mapreduce.reduce.memory.mb</code>	3754	3754
<code>yarn.app.mapreduce.am.resource.mb</code>	3754	3754

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	90112	30048
<code>yarn.nodemanager.resource.memory-mb</code>	90112	30048

## c5.18xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx1547m</code>	<code>-Xmx1547m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx3094m</code>	<code>-Xmx3094m</code>
<code>mapreduce.map.memory.mb</code>	1934	1934
<code>mapreduce.reduce.memory.mb</code>	3868	3868
<code>yarn.app.mapreduce.am.resource.mb</code>	3868	3868
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	139264	30960
<code>yarn.nodemanager.resource.memory-mb</code>	139264	30960

## c5.24xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1570m	-Xmx1570m
mapreduce.java.opts	-Xmx3140m	-Xmx3140m
mapreduce.map.memory.mb	1963	1963
mapreduce.reduce.memory.mb	3926	3926
yarn.app.mapreduce.am.resource.mb	3926	3926
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	188416	31376
yarn.nodemanager.resource.memory-mb	188416	31376

## C5a-Instances

## c5a.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1126m	-Xmx1126m
mapreduce.java.opts	-Xmx2252m	-Xmx2252m
mapreduce.map.memory.mb	1 408	1 408
mapreduce.reduce.memory.mb	2816	2816

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.app.mapreduce.am.resource.mb	2816	2816
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	5632	2816
yarn.nodemanager.resource.memory-mb	5632	2816

## c5a.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

## c5a.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx1171m</code>	<code>-Xmx1171m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.map.memory.mb</code>	1464	1464
<code>mapreduce.reduce.memory.mb</code>	2928	2928
<code>yarn.app.mapreduce.am.resource.mb</code>	2928	2928
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

## c5a.8xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx1357m</code>	<code>-Xmx1357m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2714m</code>	<code>-Xmx2714m</code>
<code>mapreduce.map.memory.mb</code>	1696	1696
<code>mapreduce.reduce.memory.mb</code>	3392	3392
<code>yarn.app.mapreduce.am.resource.mb</code>	3392	3392

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

## c5a.12xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx1502m</code>	<code>-Xmx1502m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx3004m</code>	<code>-Xmx3004m</code>
<code>mapreduce.map.memory.mb</code>	1877	1877
<code>mapreduce.reduce.memory.mb</code>	3754	3754
<code>yarn.app.mapreduce.am.resource.mb</code>	3754	3754
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	90112	30048
<code>yarn.nodemanager.resource.memory-mb</code>	90112	30048



## c5a.16xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx1459m</code>	<code>-Xmx1459m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2918m</code>	<code>-Xmx2918m</code>
<code>mapreduce.map.memory.mb</code>	1824	1824
<code>mapreduce.reduce.memory.mb</code>	3648	3648
<code>yarn.app.mapreduce.am.resource.mb</code>	3648	3648
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

## c5a.24xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx1494m</code>	<code>-Xmx1494m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2988m</code>	<code>-Xmx2988m</code>
<code>mapreduce.map.memory.mb</code>	1867	1867
<code>mapreduce.reduce.memory.mb</code>	3734	3734
<code>yarn.app.mapreduce.am.resource.mb</code>	3734	3734

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	179200	29840
<code>yarn.nodemanager.resource.memory-mb</code>	179200	29840

## C5ad-Instances

### c5ad.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx1126m</code>	<code>-Xmx1126m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2252m</code>	<code>-Xmx2252m</code>
<code>mapreduce.map.memory.mb</code>	1 408	1 408
<code>mapreduce.reduce.memory.mb</code>	2816	2816
<code>yarn.app.mapreduce.am.resource.mb</code>	2816	2816
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	5632	2816
<code>yarn.nodemanager.resource.memory-mb</code>	5632	2816

## c5ad.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx1171m</code>	<code>-Xmx1171m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.map.memory.mb</code>	1464	1464
<code>mapreduce.reduce.memory.mb</code>	2928	2928
<code>yarn.app.mapreduce.am.resource.mb</code>	2928	2928
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	11712	5856
<code>yarn.nodemanager.resource.memory-mb</code>	11712	5856

## c5ad.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx1171m</code>	<code>-Xmx1171m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.map.memory.mb</code>	1464	1464
<code>mapreduce.reduce.memory.mb</code>	2928	2928
<code>yarn.app.mapreduce.am.resource.mb</code>	2928	2928

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

## c5ad.8xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx1357m</code>	<code>-Xmx1357m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2714m</code>	<code>-Xmx2714m</code>
<code>mapreduce.map.memory.mb</code>	1696	1696
<code>mapreduce.reduce.memory.mb</code>	3392	3392
<code>yarn.app.mapreduce.am.resource.mb</code>	3392	3392
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

## c5ad.12xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1425m	-Xmx1425m
mapreduce.java.opts	-Xmx2850m	-Xmx2850m
mapreduce.map.memory.mb	1781	1781
mapreduce.reduce.memory.mb	3562	3562
yarn.app.mapreduce.am.resource.mb	3562	3562
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	85504	32074
yarn.nodemanager.resource.memory-mb	85504	32074

## c5ad.16xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1459m	-Xmx1459m
mapreduce.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.map.memory.mb	1824	1824
mapreduce.reduce.memory.mb	3648	3648
yarn.app.mapreduce.am.resource.mb	3648	3648

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

## c5ad.24xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx1494m</code>	<code>-Xmx1494m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2988m</code>	<code>-Xmx2988m</code>
<code>mapreduce.map.memory.mb</code>	1867	1867
<code>mapreduce.reduce.memory.mb</code>	3734	3734
<code>yarn.app.mapreduce.am.resource.mb</code>	3734	3734
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	179200	29840
<code>yarn.nodemanager.resource.memory-mb</code>	179200	29840

## C5d-Instances

## c5d.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1229m	-Xmx1229m
mapreduce.java.opts	-Xmx2458m	-Xmx2458m
mapreduce.map.memory.mb	1536	1536
mapreduce.reduce.memory.mb	3072	3072
yarn.app.mapreduce.am.resource.mb	3072	3072
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	6144	3072
yarn.nodemanager.resource.memory-mb	6144	3072

## c5d.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1229m	-Xmx1229m
mapreduce.java.opts	-Xmx2458m	-Xmx2458m
mapreduce.map.memory.mb	1536	1536
mapreduce.reduce.memory.mb	3072	3072

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.app.mapreduce.am.resource.mb	3072	3072
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	12288	6144
yarn.nodemanager.resource.memory-mb	12288	6144

## c5d.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1229m	-Xmx1229m
mapreduce.java.opts	-Xmx2458m	-Xmx2458m
mapreduce.map.memory.mb	1536	1536
mapreduce.reduce.memory.mb	3072	3072
yarn.app.mapreduce.am.resource.mb	3072	3072
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	24576	12288
yarn.nodemanager.resource.memory-mb	24576	12288



## c5d.9xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx1456m</code>	<code>-Xmx1456m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2912m</code>	<code>-Xmx2912m</code>
<code>mapreduce.map.memory.mb</code>	1820	1820
<code>mapreduce.reduce.memory.mb</code>	3640	3640
<code>yarn.app.mapreduce.am.resource.mb</code>	3640	3640
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	65536	32768
<code>yarn.nodemanager.resource.memory-mb</code>	65536	32768

## c5d.12xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx1502m</code>	<code>-Xmx1502m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx3004m</code>	<code>-Xmx3004m</code>
<code>mapreduce.map.memory.mb</code>	1877	1877
<code>mapreduce.reduce.memory.mb</code>	3754	3754
<code>yarn.app.mapreduce.am.resource.mb</code>	3754	3754

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	90112	30048
<code>yarn.nodemanager.resource.memory-mb</code>	90112	30048

## c5d.18xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx1547m</code>	<code>-Xmx1547m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx3094m</code>	<code>-Xmx3094m</code>
<code>mapreduce.map.memory.mb</code>	1934	1934
<code>mapreduce.reduce.memory.mb</code>	3868	3868
<code>yarn.app.mapreduce.am.resource.mb</code>	3868	3868
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	139264	30960
<code>yarn.nodemanager.resource.memory-mb</code>	139264	30960

## c5d.24xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1570m	-Xmx1570m
mapreduce.java.opts	-Xmx3140m	-Xmx3140m
mapreduce.map.memory.mb	1963	1963
mapreduce.reduce.memory.mb	3926	3926
yarn.app.mapreduce.am.resource.mb	3926	3926
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	188416	31376
yarn.nodemanager.resource.memory-mb	188416	31376

## C5n-Instances

## c5n.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1613m	-Xmx1613m
mapreduce.java.opts	-Xmx3226m	-Xmx3226m
mapreduce.map.memory.mb	2016	2016
mapreduce.reduce.memory.mb	4032	4032

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.app.mapreduce.am.resource.mb</code>	4032	4032
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	8064	4032
<code>yarn.nodemanager.resource.memory-mb</code>	8064	4032

## c5n.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx1613m</code>	<code>-Xmx1613m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx3226m</code>	<code>-Xmx3226m</code>
<code>mapreduce.map.memory.mb</code>	2016	2016
<code>mapreduce.reduce.memory.mb</code>	4032	4032
<code>yarn.app.mapreduce.am.resource.mb</code>	4032	4032
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	16128	8064
<code>yarn.nodemanager.resource.memory-mb</code>	16128	8064

## c5n.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx1741m</code>	<code>-Xmx1741m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx3482m</code>	<code>-Xmx3482m</code>
<code>mapreduce.map.memory.mb</code>	2176	2176
<code>mapreduce.reduce.memory.mb</code>	4352	4352
<code>yarn.app.mapreduce.am.resource.mb</code>	4352	4352
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	34816	17408
<code>yarn.nodemanager.resource.memory-mb</code>	34816	17408

## c5n.9xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx2002m</code>	<code>-Xmx2002m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4004m</code>	<code>-Xmx4004m</code>
<code>mapreduce.map.memory.mb</code>	2503	2503
<code>mapreduce.reduce.memory.mb</code>	5006	5006
<code>yarn.app.mapreduce.am.resource.mb</code>	5006	5006

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	90112	30040
<code>yarn.nodemanager.resource.memory-mb</code>	90112	30040

## c5n.18xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx2094m</code>	<code>-Xmx2094m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4188m</code>	<code>-Xmx4188m</code>
<code>mapreduce.map.memory.mb</code>	2617	2617
<code>mapreduce.reduce.memory.mb</code>	5234	5234
<code>yarn.app.mapreduce.am.resource.mb</code>	5234	5234
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	188416	31396
<code>yarn.nodemanager.resource.memory-mb</code>	188416	31396

## C6a-Instances

## c6a.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1126m	-Xmx1126m
mapreduce.java.opts	-Xmx2252m	-Xmx2252m
mapreduce.map.memory.mb	1 408	1 408
mapreduce.reduce.memory.mb	2816	2816
yarn.app.mapreduce.am.resource.mb	2816	2816
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	5632	2816
yarn.nodemanager.resource.memory-mb	5632	2816

## c6a.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.app.mapreduce.am.resource.mb	2928	2928
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

## c6a.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712



## c6a.8xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1357m	-Xmx1357m
mapreduce.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.map.memory.mb	1696	1696
mapreduce.reduce.memory.mb	3392	3392
yarn.app.mapreduce.am.resource.mb	3392	3392
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## c6a.12xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1425m	-Xmx1425m
mapreduce.java.opts	-Xmx2850m	-Xmx2850m
mapreduce.map.memory.mb	1781	1781
mapreduce.reduce.memory.mb	3562	3562
yarn.app.mapreduce.am.resource.mb	3562	3562

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	85504	32074
<code>yarn.nodemanager.resource.memory-mb</code>	85504	32074

## c6a.16xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx1459m</code>	<code>-Xmx1459m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2918m</code>	<code>-Xmx2918m</code>
<code>mapreduce.map.memory.mb</code>	1824	1824
<code>mapreduce.reduce.memory.mb</code>	3648	3648
<code>yarn.app.mapreduce.am.resource.mb</code>	3648	3648
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

## c6a.24xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx1494m</code>	<code>-Xmx1494m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2988m</code>	<code>-Xmx2988m</code>
<code>mapreduce.map.memory.mb</code>	1867	1867
<code>mapreduce.reduce.memory.mb</code>	3734	3734
<code>yarn.app.mapreduce.am.resource.mb</code>	3734	3734
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	179200	29840
<code>yarn.nodemanager.resource.memory-mb</code>	179200	29840

## c6a.32xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx1510m</code>	<code>-Xmx1510m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx3020m</code>	<code>-Xmx3020m</code>
<code>mapreduce.map.memory.mb</code>	1888	1888
<code>mapreduce.reduce.memory.mb</code>	3776	3776
<code>yarn.app.mapreduce.am.resource.mb</code>	3776	3776

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

## c6a.48xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx1527m</code>	<code>-Xmx1527m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx3054m</code>	<code>-Xmx3054m</code>
<code>mapreduce.map.memory.mb</code>	1909	1909
<code>mapreduce.reduce.memory.mb</code>	3818	3818
<code>yarn.app.mapreduce.am.resource.mb</code>	3818	3818
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	366592	30608
<code>yarn.nodemanager.resource.memory-mb</code>	366592	30608

## C6g-Instances

## c6g.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1126m	-Xmx1126m
mapreduce.java.opts	-Xmx2252m	-Xmx2252m
mapreduce.map.memory.mb	1 408	1 408
mapreduce.reduce.memory.mb	2816	2816
yarn.app.mapreduce.am.resource.mb	2816	2816
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	5632	2816
yarn.nodemanager.resource.memory-mb	5632	2816

## c6g.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.app.mapreduce.am.resource.mb	2928	2928
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

## c6g.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## c6g.8xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx1357m</code>	<code>-Xmx1357m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2714m</code>	<code>-Xmx2714m</code>
<code>mapreduce.map.memory.mb</code>	1696	1696
<code>mapreduce.reduce.memory.mb</code>	3392	3392
<code>yarn.app.mapreduce.am.resource.mb</code>	3392	3392
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

## c6g.12xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx1425m</code>	<code>-Xmx1425m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2850m</code>	<code>-Xmx2850m</code>
<code>mapreduce.map.memory.mb</code>	1781	1781
<code>mapreduce.reduce.memory.mb</code>	3562	3562
<code>yarn.app.mapreduce.am.resource.mb</code>	3562	3562

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	85504	32074
<code>yarn.nodemanager.resource.memory-mb</code>	85504	32074

## c6g.16xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx1459m</code>	<code>-Xmx1459m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2918m</code>	<code>-Xmx2918m</code>
<code>mapreduce.map.memory.mb</code>	1824	1824
<code>mapreduce.reduce.memory.mb</code>	3648	3648
<code>yarn.app.mapreduce.am.resource.mb</code>	3648	3648
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184



## C6gd-Instances

## c6gd.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1126m	-Xmx1126m
mapreduce.java.opts	-Xmx2252m	-Xmx2252m
mapreduce.map.memory.mb	1 408	1 408
mapreduce.reduce.memory.mb	2816	2816
yarn.app.mapreduce.am.resource.mb	2816	2816
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	5632	2816
yarn.nodemanager.resource.memory-mb	5632	2816

## c6gd.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.app.mapreduce.am.resource.mb	2928	2928
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

## c6gd.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## c6gd.8xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1357m	-Xmx1357m
mapreduce.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.map.memory.mb	1696	1696
mapreduce.reduce.memory.mb	3392	3392
yarn.app.mapreduce.am.resource.mb	3392	3392
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## c6gd.12xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1425m	-Xmx1425m
mapreduce.java.opts	-Xmx2850m	-Xmx2850m
mapreduce.map.memory.mb	1781	1781
mapreduce.reduce.memory.mb	3562	3562
yarn.app.mapreduce.am.resource.mb	3562	3562

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	85504	32074
<code>yarn.nodemanager.resource.memory-mb</code>	85504	32074

## c6gd.16xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx1459m</code>	<code>-Xmx1459m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2918m</code>	<code>-Xmx2918m</code>
<code>mapreduce.map.memory.mb</code>	1824	1824
<code>mapreduce.reduce.memory.mb</code>	3648	3648
<code>yarn.app.mapreduce.am.resource.mb</code>	3648	3648
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

## C6gn-Instances

## c6gn.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1126m	-Xmx1126m
mapreduce.java.opts	-Xmx2252m	-Xmx2252m
mapreduce.map.memory.mb	1 408	1 408
mapreduce.reduce.memory.mb	2816	2816
yarn.app.mapreduce.am.resource.mb	2816	2816
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	5632	2816
yarn.nodemanager.resource.memory-mb	5632	2816

## c6gn.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.app.mapreduce.am.resource.mb	2928	2928
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

## c6gn.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## c6gn.8xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1357m	-Xmx1357m
mapreduce.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.map.memory.mb	1696	1696
mapreduce.reduce.memory.mb	3392	3392
yarn.app.mapreduce.am.resource.mb	3392	3392
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## c6gn.12xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1425m	-Xmx1425m
mapreduce.java.opts	-Xmx2850m	-Xmx2850m
mapreduce.map.memory.mb	1781	1781
mapreduce.reduce.memory.mb	3562	3562
yarn.app.mapreduce.am.resource.mb	3562	3562

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	85504	32074
<code>yarn.nodemanager.resource.memory-mb</code>	85504	32074

## c6gn.16xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx1459m</code>	<code>-Xmx1459m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2918m</code>	<code>-Xmx2918m</code>
<code>mapreduce.map.memory.mb</code>	1824	1824
<code>mapreduce.reduce.memory.mb</code>	3648	3648
<code>yarn.app.mapreduce.am.resource.mb</code>	3648	3648
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184



## C6i-Instances

## c6i.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1126m	-Xmx1126m
mapreduce.java.opts	-Xmx2252m	-Xmx2252m
mapreduce.map.memory.mb	1 408	1 408
mapreduce.reduce.memory.mb	2816	2816
yarn.app.mapreduce.am.resource.mb	2816	2816
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	5632	2816
yarn.nodemanager.resource.memory-mb	5632	2816

## c6i.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.app.mapreduce.am.resource.mb	2928	2928
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

## c6i.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## c6i.8xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1357m	-Xmx1357m
mapreduce.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.map.memory.mb	1696	1696
mapreduce.reduce.memory.mb	3392	3392
yarn.app.mapreduce.am.resource.mb	3392	3392
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## c6i.12xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1425m	-Xmx1425m
mapreduce.java.opts	-Xmx2850m	-Xmx2850m
mapreduce.map.memory.mb	1781	1781
mapreduce.reduce.memory.mb	3562	3562
yarn.app.mapreduce.am.resource.mb	3562	3562

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	85504	32074
<code>yarn.nodemanager.resource.memory-mb</code>	85504	32074

## c6i.16xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx1459m</code>	<code>-Xmx1459m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2918m</code>	<code>-Xmx2918m</code>
<code>mapreduce.map.memory.mb</code>	1824	1824
<code>mapreduce.reduce.memory.mb</code>	3648	3648
<code>yarn.app.mapreduce.am.resource.mb</code>	3648	3648
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

## c6i.24xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1494m	-Xmx1494m
mapreduce.java.opts	-Xmx2988m	-Xmx2988m
mapreduce.map.memory.mb	1867	1867
mapreduce.reduce.memory.mb	3734	3734
yarn.app.mapreduce.am.resource.mb	3734	3734
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	179200	29840
yarn.nodemanager.resource.memory-mb	179200	29840

## c6i.32xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1510m	-Xmx1510m
mapreduce.java.opts	-Xmx3020m	-Xmx3020m
mapreduce.map.memory.mb	1888	1888
mapreduce.reduce.memory.mb	3776	3776
yarn.app.mapreduce.am.resource.mb	3776	3776

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

## C6id-Instances

### c6id.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx1126m</code>	<code>-Xmx1126m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2252m</code>	<code>-Xmx2252m</code>
<code>mapreduce.map.memory.mb</code>	1 408	1 408
<code>mapreduce.reduce.memory.mb</code>	2816	2816
<code>yarn.app.mapreduce.am.resource.mb</code>	2816	2816
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	5632	2816
<code>yarn.nodemanager.resource.memory-mb</code>	5632	2816

## c6id.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx1171m</code>	<code>-Xmx1171m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.map.memory.mb</code>	1464	1464
<code>mapreduce.reduce.memory.mb</code>	2928	2928
<code>yarn.app.mapreduce.am.resource.mb</code>	2928	2928
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	11712	5856
<code>yarn.nodemanager.resource.memory-mb</code>	11712	5856

## c6id.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx1171m</code>	<code>-Xmx1171m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.map.memory.mb</code>	1464	1464
<code>mapreduce.reduce.memory.mb</code>	2928	2928
<code>yarn.app.mapreduce.am.resource.mb</code>	2928	2928

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

## c6id.8xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx1357m</code>	<code>-Xmx1357m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2714m</code>	<code>-Xmx2714m</code>
<code>mapreduce.map.memory.mb</code>	1696	1696
<code>mapreduce.reduce.memory.mb</code>	3392	3392
<code>yarn.app.mapreduce.am.resource.mb</code>	3392	3392
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136



## c6id.12xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx1425m</code>	<code>-Xmx1425m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2850m</code>	<code>-Xmx2850m</code>
<code>mapreduce.map.memory.mb</code>	1781	1781
<code>mapreduce.reduce.memory.mb</code>	3562	3562
<code>yarn.app.mapreduce.am.resource.mb</code>	3562	3562
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	85504	32074
<code>yarn.nodemanager.resource.memory-mb</code>	85504	32074

## c6id.16xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx1459m</code>	<code>-Xmx1459m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2918m</code>	<code>-Xmx2918m</code>
<code>mapreduce.map.memory.mb</code>	1824	1824
<code>mapreduce.reduce.memory.mb</code>	3648	3648
<code>yarn.app.mapreduce.am.resource.mb</code>	3648	3648

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

## c6id.24xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx1494m</code>	<code>-Xmx1494m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2988m</code>	<code>-Xmx2988m</code>
<code>mapreduce.map.memory.mb</code>	1867	1867
<code>mapreduce.reduce.memory.mb</code>	3734	3734
<code>yarn.app.mapreduce.am.resource.mb</code>	3734	3734
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	179200	29840
<code>yarn.nodemanager.resource.memory-mb</code>	179200	29840

## c6id.32xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1510m	-Xmx1510m
mapreduce.java.opts	-Xmx3020m	-Xmx3020m
mapreduce.map.memory.mb	1888	1888
mapreduce.reduce.memory.mb	3776	3776
yarn.app.mapreduce.am.resource.mb	3776	3776
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

## C6in-Instances

## c6in.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1126m	-Xmx1126m
mapreduce.java.opts	-Xmx2252m	-Xmx2252m
mapreduce.map.memory.mb	1 408	1 408
mapreduce.reduce.memory.mb	2816	2816

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.app.mapreduce.am.resource.mb	2816	2816
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	5632	2816
yarn.nodemanager.resource.memory-mb	5632	2816

## c6in.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

## c6in.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## c6in.8xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1357m	-Xmx1357m
mapreduce.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.map.memory.mb	1696	1696
mapreduce.reduce.memory.mb	3392	3392
yarn.app.mapreduce.am.resource.mb	3392	3392

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

## c6in.12xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx1425m</code>	<code>-Xmx1425m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2850m</code>	<code>-Xmx2850m</code>
<code>mapreduce.map.memory.mb</code>	1781	1781
<code>mapreduce.reduce.memory.mb</code>	3562	3562
<code>yarn.app.mapreduce.am.resource.mb</code>	3562	3562
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	85504	32074
<code>yarn.nodemanager.resource.memory-mb</code>	85504	32074

## c6in.16xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1459m	-Xmx1459m
mapreduce.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.map.memory.mb	1824	1824
mapreduce.reduce.memory.mb	3648	3648
yarn.app.mapreduce.am.resource.mb	3648	3648
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## c6in.24xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1494m	-Xmx1494m
mapreduce.java.opts	-Xmx2988m	-Xmx2988m
mapreduce.map.memory.mb	1867	1867
mapreduce.reduce.memory.mb	3734	3734
yarn.app.mapreduce.am.resource.mb	3734	3734

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	179200	29840
<code>yarn.nodemanager.resource.memory-mb</code>	179200	29840

## c6in.32xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx1510m</code>	<code>-Xmx1510m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx3020m</code>	<code>-Xmx3020m</code>
<code>mapreduce.map.memory.mb</code>	1888	1888
<code>mapreduce.reduce.memory.mb</code>	3776	3776
<code>yarn.app.mapreduce.am.resource.mb</code>	3776	3776
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208



## c7a-Instances

## c7a.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1126m	-Xmx1126m
mapreduce.java.opts	-Xmx2252m	-Xmx2252m
mapreduce.map.memory.mb	1 408	1 408
mapreduce.reduce.memory.mb	2816	2816
yarn.app.mapreduce.am.resource.mb	2816	2816
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	5632	2816
yarn.nodemanager.resource.memory-mb	5632	2816

## c7a.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.app.mapreduce.am.resource.mb	2928	2928
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

## c7a.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## c7a.8xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx1357m</code>	<code>-Xmx1357m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2714m</code>	<code>-Xmx2714m</code>
<code>mapreduce.map.memory.mb</code>	1696	1696
<code>mapreduce.reduce.memory.mb</code>	3392	3392
<code>yarn.app.mapreduce.am.resource.mb</code>	3392	3392
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

## c7a.12xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx1425m</code>	<code>-Xmx1425m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2850m</code>	<code>-Xmx2850m</code>
<code>mapreduce.map.memory.mb</code>	1781	1781
<code>mapreduce.reduce.memory.mb</code>	3562	3562
<code>yarn.app.mapreduce.am.resource.mb</code>	3562	3562

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	85504	32074
<code>yarn.nodemanager.resource.memory-mb</code>	85504	32074

## c7a.16xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx1459m</code>	<code>-Xmx1459m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2918m</code>	<code>-Xmx2918m</code>
<code>mapreduce.map.memory.mb</code>	1824	1824
<code>mapreduce.reduce.memory.mb</code>	3648	3648
<code>yarn.app.mapreduce.am.resource.mb</code>	3648	3648
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

## c7a.24xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1494m	-Xmx1494m
mapreduce.java.opts	-Xmx2988m	-Xmx2988m
mapreduce.map.memory.mb	1867	1867
mapreduce.reduce.memory.mb	3734	3734
yarn.app.mapreduce.am.resource.mb	3734	3734
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	179200	29840
yarn.nodemanager.resource.memory-mb	179200	29840

## c7a.32xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1510m	-Xmx1510m
mapreduce.java.opts	-Xmx3020m	-Xmx3020m
mapreduce.map.memory.mb	1888	1888
mapreduce.reduce.memory.mb	3776	3776
yarn.app.mapreduce.am.resource.mb	3776	3776

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

## c7a.48xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx1527m</code>	<code>-Xmx1527m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx3054m</code>	<code>-Xmx3054m</code>
<code>mapreduce.map.memory.mb</code>	1909	1909
<code>mapreduce.reduce.memory.mb</code>	3818	3818
<code>yarn.app.mapreduce.am.resource.mb</code>	3818	3818
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	366592	30608
<code>yarn.nodemanager.resource.memory-mb</code>	366592	30608

## C7g-Instances

## c7g.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1126m	-Xmx1126m
mapreduce.java.opts	-Xmx2252m	-Xmx2252m
mapreduce.map.memory.mb	1 408	1 408
mapreduce.reduce.memory.mb	2816	2816
yarn.app.mapreduce.am.resource.mb	2816	2816
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	5632	2816
yarn.nodemanager.resource.memory-mb	5632	2816

## c7g.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.app.mapreduce.am.resource.mb	2928	2928
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

## c7g.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712



## c7g.8xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx1357m</code>	<code>-Xmx1357m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2714m</code>	<code>-Xmx2714m</code>
<code>mapreduce.map.memory.mb</code>	1696	1696
<code>mapreduce.reduce.memory.mb</code>	3392	3392
<code>yarn.app.mapreduce.am.resource.mb</code>	3392	3392
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

## c7g.12xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx1425m</code>	<code>-Xmx1425m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2850m</code>	<code>-Xmx2850m</code>
<code>mapreduce.map.memory.mb</code>	1781	1781
<code>mapreduce.reduce.memory.mb</code>	3562	3562
<code>yarn.app.mapreduce.am.resource.mb</code>	3562	3562

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	85504	32074
yarn.nodemanager.resource.memory-mb	85504	32074

## c7g.16xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1459m	-Xmx1459m
mapreduce.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.map.memory.mb	1824	1824
mapreduce.reduce.memory.mb	3648	3648
yarn.app.mapreduce.am.resource.mb	3648	3648
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## C7gd-Instances

## c7gd.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1126m	-Xmx1126m
mapreduce.java.opts	-Xmx2252m	-Xmx2252m
mapreduce.map.memory.mb	1 408	1 408
mapreduce.reduce.memory.mb	2816	2816
yarn.app.mapreduce.am.resource.mb	2816	2816
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	5632	2816
yarn.nodemanager.resource.memory-mb	5632	2816

## c7gd.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.app.mapreduce.am.resource.mb	2928	2928
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

## c7gd.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## c7gd.8xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1357m	-Xmx1357m
mapreduce.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.map.memory.mb	1696	1696
mapreduce.reduce.memory.mb	3392	3392
yarn.app.mapreduce.am.resource.mb	3392	3392
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## c7gd.12xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1425m	-Xmx1425m
mapreduce.java.opts	-Xmx2850m	-Xmx2850m
mapreduce.map.memory.mb	1781	1781
mapreduce.reduce.memory.mb	3562	3562
yarn.app.mapreduce.am.resource.mb	3562	3562

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	85504	32074
<code>yarn.nodemanager.resource.memory-mb</code>	85504	32074

## c7gd.16xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx1459m</code>	<code>-Xmx1459m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2918m</code>	<code>-Xmx2918m</code>
<code>mapreduce.map.memory.mb</code>	1824	1824
<code>mapreduce.reduce.memory.mb</code>	3648	3648
<code>yarn.app.mapreduce.am.resource.mb</code>	3648	3648
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

## C7gn-Instances

## c7gn.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1126m	-Xmx1126m
mapreduce.java.opts	-Xmx2252m	-Xmx2252m
mapreduce.map.memory.mb	1 408	1 408
mapreduce.reduce.memory.mb	2816	2816
yarn.app.mapreduce.am.resource.mb	2816	2816
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	5632	2816
yarn.nodemanager.resource.memory-mb	5632	2816

## c7gn.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.app.mapreduce.am.resource.mb	2928	2928
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

## c7gn.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712



## c7gn.8xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1357m	-Xmx1357m
mapreduce.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.map.memory.mb	1696	1696
mapreduce.reduce.memory.mb	3392	3392
yarn.app.mapreduce.am.resource.mb	3392	3392
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## c7gn.12xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1425m	-Xmx1425m
mapreduce.java.opts	-Xmx2850m	-Xmx2850m
mapreduce.map.memory.mb	1781	1781
mapreduce.reduce.memory.mb	3562	3562
yarn.app.mapreduce.am.resource.mb	3562	3562

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	85504	32074
<code>yarn.nodemanager.resource.memory-mb</code>	85504	32074

## c7gn.16xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx1459m</code>	<code>-Xmx1459m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2918m</code>	<code>-Xmx2918m</code>
<code>mapreduce.map.memory.mb</code>	1824	1824
<code>mapreduce.reduce.memory.mb</code>	3648	3648
<code>yarn.app.mapreduce.am.resource.mb</code>	3648	3648
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

## C7i-Instances

## c7i.x groß

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1126m	-Xmx1126m
mapreduce.java.opts	-Xmx2252m	-Xmx2252m
mapreduce.map.memory.mb	1 408	1 408
mapreduce.reduce.memory.mb	2816	2816
yarn.app.mapreduce.am.resource.mb	2816	2816
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	5632	2816
yarn.nodemanager.resource.memory-mb	5632	2816

## c7i.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.app.mapreduce.am.resource.mb	2928	2928
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

## c7i.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## c7i.8xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx1357m</code>	<code>-Xmx1357m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2714m</code>	<code>-Xmx2714m</code>
<code>mapreduce.map.memory.mb</code>	1696	1696
<code>mapreduce.reduce.memory.mb</code>	3392	3392
<code>yarn.app.mapreduce.am.resource.mb</code>	3392	3392
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

## c7i.12xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx1425m</code>	<code>-Xmx1425m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2850m</code>	<code>-Xmx2850m</code>
<code>mapreduce.map.memory.mb</code>	1781	1781
<code>mapreduce.reduce.memory.mb</code>	3562	3562
<code>yarn.app.mapreduce.am.resource.mb</code>	3562	3562

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	85504	32074
<code>yarn.nodemanager.resource.memory-mb</code>	85504	32074

## c7i.16xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx1459m</code>	<code>-Xmx1459m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2918m</code>	<code>-Xmx2918m</code>
<code>mapreduce.map.memory.mb</code>	1824	1824
<code>mapreduce.reduce.memory.mb</code>	3648	3648
<code>yarn.app.mapreduce.am.resource.mb</code>	3648	3648
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

## c7i.24xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1494m	-Xmx1494m
mapreduce.java.opts	-Xmx2988m	-Xmx2988m
mapreduce.map.memory.mb	1867	1867
mapreduce.reduce.memory.mb	3734	3734
yarn.app.mapreduce.am.resource.mb	3734	3734
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	179200	29840
yarn.nodemanager.resource.memory-mb	179200	29840

## c7i.48xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1527m	-Xmx1527m
mapreduce.java.opts	-Xmx3054m	-Xmx3054m
mapreduce.map.memory.mb	1909	1909
mapreduce.reduce.memory.mb	3818	3818
yarn.app.mapreduce.am.resource.mb	3818	3818

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	366592	30608
<code>yarn.nodemanager.resource.memory-mb</code>	366592	30608

## D2-Instances

### d2.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4684m</code>	<code>-Xmx4684m</code>
<code>mapreduce.map.memory.mb</code>	2928	2928
<code>mapreduce.reduce.memory.mb</code>	5856	5856
<code>yarn.app.mapreduce.am.resource.mb</code>	5856	5856
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712



## d2.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx2714m</code>	<code>-Xmx2714m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5428m</code>	<code>-Xmx5428m</code>
<code>mapreduce.map.memory.mb</code>	3392	3392
<code>mapreduce.reduce.memory.mb</code>	6784	6784
<code>yarn.app.mapreduce.am.resource.mb</code>	6784	6784
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

## d2.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx2918m</code>	<code>-Xmx2918m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5836m</code>	<code>-Xmx5836m</code>
<code>mapreduce.map.memory.mb</code>	3648	3648
<code>mapreduce.reduce.memory.mb</code>	7296	7296
<code>yarn.app.mapreduce.am.resource.mb</code>	7296	7296

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

## d2.8xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx2417m</code>	<code>-Xmx2417m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4834m</code>	<code>-Xmx4834m</code>
<code>mapreduce.map.memory.mb</code>	3021	3021
<code>mapreduce.reduce.memory.mb</code>	6042	6042
<code>yarn.app.mapreduce.am.resource.mb</code>	6042	6042
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30194
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30194

## D3-Instances

## d3.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx4685m	-Xmx4685m
mapreduce.java.opts	-Xmx9370m	-Xmx9370m
mapreduce.map.memory.mb	5856	5856
mapreduce.reduce.memory.mb	11712	11712
yarn.app.mapreduce.am.resource.mb	11712	11712
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## d3.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx5427m	-Xmx5427m
mapreduce.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.map.memory.mb	6784	6784
mapreduce.reduce.memory.mb	13568	13568

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.app.mapreduce.am.resource.mb	13568	13568
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## d3.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx5837m	-Xmx5837m
mapreduce.java.opts	-Xmx11674m	-Xmx11674m
mapreduce.map.memory.mb	7296	7296
mapreduce.reduce.memory.mb	14592	14592
yarn.app.mapreduce.am.resource.mb	14592	14592
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## d3.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.java.opts	-Xmx12084m	-Xmx12084m
mapreduce.map.memory.mb	7552	7552
mapreduce.reduce.memory.mb	15104	15104
yarn.app.mapreduce.am.resource.mb	15104	15104
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

## D3en-Instances

## d3en.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

## d3en.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## d3en.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx2714m</code>	<code>-Xmx2714m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5428m</code>	<code>-Xmx5428m</code>
<code>mapreduce.map.memory.mb</code>	3392	3392
<code>mapreduce.reduce.memory.mb</code>	6784	6784
<code>yarn.app.mapreduce.am.resource.mb</code>	6784	6784
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

## d3en.6xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx2850m</code>	<code>-Xmx2850m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5700m</code>	<code>-Xmx5700m</code>
<code>mapreduce.map.memory.mb</code>	3563	3563
<code>mapreduce.reduce.memory.mb</code>	7126	7126
<code>yarn.app.mapreduce.am.resource.mb</code>	7126	7126

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	85504	28496
<code>yarn.nodemanager.resource.memory-mb</code>	85504	28496

## d3en.8xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx2918m</code>	<code>-Xmx2918m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5836m</code>	<code>-Xmx5836m</code>
<code>mapreduce.map.memory.mb</code>	3648	3648
<code>mapreduce.reduce.memory.mb</code>	7296	7296
<code>yarn.app.mapreduce.am.resource.mb</code>	7296	7296
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184



## d3en.12xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx2986m	-Xmx2986m
mapreduce.java.opts	-Xmx5972m	-Xmx5972m
mapreduce.map.memory.mb	3733	3733
mapreduce.reduce.memory.mb	7466	7466
yarn.app.mapreduce.am.resource.mb	7466	7466
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	179200	29880
yarn.nodemanager.resource.memory-mb	179200	29880

## G3-Instances

## g3.4xgroß

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx5837m	-Xmx5837m
mapreduce.java.opts	-Xmx11674m	-Xmx11674m
mapreduce.map.memory.mb	7296	7296
mapreduce.reduce.memory.mb	14592	14592

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.app.mapreduce.am.resource.mb	14592	14592
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## g3.8xgroß

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.java.opts	-Xmx12084m	-Xmx12084m
mapreduce.map.memory.mb	7552	7552
mapreduce.reduce.memory.mb	15104	15104
yarn.app.mapreduce.am.resource.mb	15104	15104
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

## g3.16xgroß

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6144m</code>	<code>-Xmx6144m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12288m</code>	<code>-Xmx12288m</code>
<code>mapreduce.map.memory.mb</code>	7680	7680
<code>mapreduce.reduce.memory.mb</code>	15360	15360
<code>yarn.app.mapreduce.am.resource.mb</code>	15360	15360
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	491520	30720
<code>yarn.nodemanager.resource.memory-mb</code>	491520	30720

## G3s-Instances

## g3s.xgroß

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx4685m</code>	<code>-Xmx4685m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx9370m</code>	<code>-Xmx9370m</code>
<code>mapreduce.map.memory.mb</code>	5856	5856
<code>mapreduce.reduce.memory.mb</code>	11712	11712

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.app.mapreduce.am.resource.mb	11712	11712
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## G4dn-Instances

### g4dn.xgroß

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx2458m	-Xmx2458m
mapreduce.java.opts	-Xmx4916m	-Xmx4916m
mapreduce.map.memory.mb	3072	3072
mapreduce.reduce.memory.mb	6144	6144
yarn.app.mapreduce.am.resource.mb	6144	6144
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	12288	6144

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.nodemanager.resource.memory-mb</code>	12288	6144

## g4dn.2xgroß

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx2458m</code>	<code>-Xmx2458m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4916m</code>	<code>-Xmx4916m</code>
<code>mapreduce.map.memory.mb</code>	3072	3072
<code>mapreduce.reduce.memory.mb</code>	6144	6144
<code>yarn.app.mapreduce.am.resource.mb</code>	6144	6144
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	24576	12288
<code>yarn.nodemanager.resource.memory-mb</code>	24576	12288

## g4dn.4xgroß

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx2867m</code>	<code>-Xmx2867m</code>

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.java.opts	-Xmx5734m	-Xmx5734m
mapreduce.map.memory.mb	3584	3584
mapreduce.reduce.memory.mb	7168	7168
yarn.app.mapreduce.am.resource.mb	7168	7168
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	57344	28672
yarn.nodemanager.resource.memory-mb	57344	28672

## g4dn.8xgroß

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx3072m	-Xmx3072m
mapreduce.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.map.memory.mb	3840	3840
mapreduce.reduce.memory.mb	7680	7680
yarn.app.mapreduce.am.resource.mb	7680	7680
yarn.scheduler.minimum-allocation-mb	32	32

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.scheduler.maximum-allocation-mb	122880	30720
yarn.nodemanager.resource.memory-mb	122880	30720

## g4dn.12xgroß

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx3140m	-Xmx3140m
mapreduce.java.opts	-Xmx6280m	-Xmx6280m
mapreduce.map.memory.mb	3925	3925
mapreduce.reduce.memory.mb	7850	7850
yarn.app.mapreduce.am.resource.mb	7850	7850
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	188416	31416
yarn.nodemanager.resource.memory-mb	188416	31416

## g4dn.16xgroß

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx3174m</code>	<code>-Xmx3174m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6348m</code>	<code>-Xmx6348m</code>
<code>mapreduce.map.memory.mb</code>	3968	3968
<code>mapreduce.reduce.memory.mb</code>	7936	7936
<code>yarn.app.mapreduce.am.resource.mb</code>	7936	7936
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	253952	31744
<code>yarn.nodemanager.resource.memory-mb</code>	253952	31744

## G5-Instances

## g5.xgroß

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4684m</code>	<code>-Xmx4684m</code>
<code>mapreduce.map.memory.mb</code>	2928	2928
<code>mapreduce.reduce.memory.mb</code>	5856	5856



Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

## g5.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## g5.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx2714m</code>	<code>-Xmx2714m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5428m</code>	<code>-Xmx5428m</code>
<code>mapreduce.map.memory.mb</code>	3392	3392
<code>mapreduce.reduce.memory.mb</code>	6784	6784
<code>yarn.app.mapreduce.am.resource.mb</code>	6784	6784
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

## g5.8xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx2918m</code>	<code>-Xmx2918m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5836m</code>	<code>-Xmx5836m</code>
<code>mapreduce.map.memory.mb</code>	3648	3648
<code>mapreduce.reduce.memory.mb</code>	7296	7296
<code>yarn.app.mapreduce.am.resource.mb</code>	7296	7296

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

## g5.12xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx2986m</code>	<code>-Xmx2986m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5972m</code>	<code>-Xmx5972m</code>
<code>mapreduce.map.memory.mb</code>	3733	3733
<code>mapreduce.reduce.memory.mb</code>	7466	7466
<code>yarn.app.mapreduce.am.resource.mb</code>	7466	7466
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	179200	29880
<code>yarn.nodemanager.resource.memory-mb</code>	179200	29880

## g5.16xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx3021m</code>	<code>-Xmx3021m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.map.memory.mb</code>	3776	3776
<code>mapreduce.reduce.memory.mb</code>	7552	7552
<code>yarn.app.mapreduce.am.resource.mb</code>	7552	7552
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

## g5.24xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx3055m</code>	<code>-Xmx3055m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6110m</code>	<code>-Xmx6110m</code>
<code>mapreduce.map.memory.mb</code>	3819	3819
<code>mapreduce.reduce.memory.mb</code>	7638	7638
<code>yarn.app.mapreduce.am.resource.mb</code>	7638	7638

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	366592	30520
<code>yarn.nodemanager.resource.memory-mb</code>	366592	30520

## g5.48xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx3089m</code>	<code>-Xmx3089m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6178m</code>	<code>-Xmx6178m</code>
<code>mapreduce.map.memory.mb</code>	3861	3861
<code>mapreduce.reduce.memory.mb</code>	7722	7722
<code>yarn.app.mapreduce.am.resource.mb</code>	7722	7722
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	741376	30952
<code>yarn.nodemanager.resource.memory-mb</code>	741376	30952

## H1-Instances

## h1.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx2458m	-Xmx2458m
mapreduce.java.opts	-Xmx4916m	-Xmx4916m
mapreduce.map.memory.mb	3072	3072
mapreduce.reduce.memory.mb	6144	6144
yarn.app.mapreduce.am.resource.mb	6144	6144
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	24576	12288
yarn.nodemanager.resource.memory-mb	24576	12288

## h1.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx2867m	-Xmx2867m
mapreduce.java.opts	-Xmx5734m	-Xmx5734m
mapreduce.map.memory.mb	3584	3584
mapreduce.reduce.memory.mb	7168	7168

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.app.mapreduce.am.resource.mb	7168	7168
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	57344	28672
yarn.nodemanager.resource.memory-mb	57344	28672

## h1.8xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx3072m	-Xmx3072m
mapreduce.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.map.memory.mb	3840	3840
mapreduce.reduce.memory.mb	7680	7680
yarn.app.mapreduce.am.resource.mb	7680	7680
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	122880	30720
yarn.nodemanager.resource.memory-mb	122880	30720

## h1.16xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx3174m</code>	<code>-Xmx3174m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6348m</code>	<code>-Xmx6348m</code>
<code>mapreduce.map.memory.mb</code>	3968	3968
<code>mapreduce.reduce.memory.mb</code>	7936	7936
<code>yarn.app.mapreduce.am.resource.mb</code>	7936	7936
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	253952	31744
<code>yarn.nodemanager.resource.memory-mb</code>	253952	31744

## I2-Instances

## i2.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4684m</code>	<code>-Xmx4684m</code>
<code>mapreduce.map.memory.mb</code>	2928	2928
<code>mapreduce.reduce.memory.mb</code>	5856	5856



Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## i2.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.java.opts	-Xmx5428m	-Xmx5428m
mapreduce.map.memory.mb	3392	3392
mapreduce.reduce.memory.mb	6784	6784
yarn.app.mapreduce.am.resource.mb	6784	6784
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## i2.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.java.opts	-Xmx5836m	-Xmx5836m
mapreduce.map.memory.mb	3648	3648
mapreduce.reduce.memory.mb	7296	7296
yarn.app.mapreduce.am.resource.mb	7296	7296
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## i2.8xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx3021m	-Xmx3021m
mapreduce.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.map.memory.mb	3776	3776
mapreduce.reduce.memory.mb	7552	7552
yarn.app.mapreduce.am.resource.mb	7552	7552

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

### I3-Instances

#### i3.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx4685m</code>	<code>-Xmx4685m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx9370m</code>	<code>-Xmx9370m</code>
<code>mapreduce.map.memory.mb</code>	5856	5856
<code>mapreduce.reduce.memory.mb</code>	11712	11712
<code>yarn.app.mapreduce.am.resource.mb</code>	11712	11712
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

## i3.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx5427m</code>	<code>-Xmx5427m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx10854m</code>	<code>-Xmx10854m</code>
<code>mapreduce.map.memory.mb</code>	6784	6784
<code>mapreduce.reduce.memory.mb</code>	13568	13568
<code>yarn.app.mapreduce.am.resource.mb</code>	13568	13568
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

## i3.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx5837m</code>	<code>-Xmx5837m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx11674m</code>	<code>-Xmx11674m</code>
<code>mapreduce.map.memory.mb</code>	7296	7296
<code>mapreduce.reduce.memory.mb</code>	14592	14592
<code>yarn.app.mapreduce.am.resource.mb</code>	14592	14592

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## i3.8xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.java.opts	-Xmx12084m	-Xmx12084m
mapreduce.map.memory.mb	7552	7552
mapreduce.reduce.memory.mb	15104	15104
yarn.app.mapreduce.am.resource.mb	15104	15104
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

## i3.16xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6144m</code>	<code>-Xmx6144m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12288m</code>	<code>-Xmx12288m</code>
<code>mapreduce.map.memory.mb</code>	7680	7680
<code>mapreduce.reduce.memory.mb</code>	15360	15360
<code>yarn.app.mapreduce.am.resource.mb</code>	15360	15360
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	491520	30720
<code>yarn.nodemanager.resource.memory-mb</code>	491520	30720

## I3en-Instances

## i3en.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx4915m</code>	<code>-Xmx4915m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx9830m</code>	<code>-Xmx9830m</code>
<code>mapreduce.map.memory.mb</code>	6144	6144
<code>mapreduce.reduce.memory.mb</code>	12288	12288

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.app.mapreduce.am.resource.mb	12288	12288
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	24576	12288
yarn.nodemanager.resource.memory-mb	24576	12288

## i3en.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx5734m	-Xmx5734m
mapreduce.java.opts	-Xmx11468m	-Xmx11468m
mapreduce.map.memory.mb	7168	7168
mapreduce.reduce.memory.mb	14336	14336
yarn.app.mapreduce.am.resource.mb	14336	14336
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	57344	28672
yarn.nodemanager.resource.memory-mb	57344	28672

## i3en.3xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6007m</code>	<code>-Xmx6007m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12014m</code>	<code>-Xmx12014m</code>
<code>mapreduce.map.memory.mb</code>	7509	7509
<code>mapreduce.reduce.memory.mb</code>	15018	15018
<code>yarn.app.mapreduce.am.resource.mb</code>	15018	15018
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	90112	30040
<code>yarn.nodemanager.resource.memory-mb</code>	90112	30040

## i3en.6xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6281m</code>	<code>-Xmx6281m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12562m</code>	<code>-Xmx12562m</code>
<code>mapreduce.map.memory.mb</code>	7851	7851
<code>mapreduce.reduce.memory.mb</code>	15702	15702
<code>yarn.app.mapreduce.am.resource.mb</code>	15702	15702



Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	188416	31396
<code>yarn.nodemanager.resource.memory-mb</code>	188416	31396

## i3en.12xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6417m</code>	<code>-Xmx6417m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12834m</code>	<code>-Xmx12834m</code>
<code>mapreduce.map.memory.mb</code>	8021	8021
<code>mapreduce.reduce.memory.mb</code>	16042	16042
<code>yarn.app.mapreduce.am.resource.mb</code>	16042	16042
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	385024	32100
<code>yarn.nodemanager.resource.memory-mb</code>	385024	32100

## i3en.24xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx6486m	-Xmx6486m
mapreduce.java.opts	-Xmx12972m	-Xmx12972m
mapreduce.map.memory.mb	8107	8107
mapreduce.reduce.memory.mb	16214	16214
yarn.app.mapreduce.am.resource.mb	16214	16214
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	778240	32396
yarn.nodemanager.resource.memory-mb	778240	32396

## l4g-Instances

## i4g.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx4685m	-Xmx4685m
mapreduce.java.opts	-Xmx9370m	-Xmx9370m
mapreduce.map.memory.mb	5856	5856
mapreduce.reduce.memory.mb	11712	11712

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.app.mapreduce.am.resource.mb	11712	11712
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## i4g.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx5427m	-Xmx5427m
mapreduce.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.map.memory.mb	6784	6784
mapreduce.reduce.memory.mb	13568	13568
yarn.app.mapreduce.am.resource.mb	13568	13568
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## i4g.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx5837m</code>	<code>-Xmx5837m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx11674m</code>	<code>-Xmx11674m</code>
<code>mapreduce.map.memory.mb</code>	7296	7296
<code>mapreduce.reduce.memory.mb</code>	14592	14592
<code>yarn.app.mapreduce.am.resource.mb</code>	14592	14592
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

## i4g.8xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12084m</code>	<code>-Xmx12084m</code>
<code>mapreduce.map.memory.mb</code>	7552	7552
<code>mapreduce.reduce.memory.mb</code>	15104	15104
<code>yarn.app.mapreduce.am.resource.mb</code>	15104	15104

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

## i4g.16xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6144m</code>	<code>-Xmx6144m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12288m</code>	<code>-Xmx12288m</code>
<code>mapreduce.map.memory.mb</code>	7680	7680
<code>mapreduce.reduce.memory.mb</code>	15360	15360
<code>yarn.app.mapreduce.am.resource.mb</code>	15360	15360
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	491520	30720
<code>yarn.nodemanager.resource.memory-mb</code>	491520	30720

## I4i-Instances

## i4i.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx4685m	-Xmx4685m
mapreduce.java.opts	-Xmx9370m	-Xmx9370m
mapreduce.map.memory.mb	5856	5856
mapreduce.reduce.memory.mb	11712	11712
yarn.app.mapreduce.am.resource.mb	11712	11712
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## i4i.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx5427m	-Xmx5427m
mapreduce.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.map.memory.mb	6784	6784
mapreduce.reduce.memory.mb	13568	13568

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.app.mapreduce.am.resource.mb	13568	13568
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## i4i.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx5837m	-Xmx5837m
mapreduce.java.opts	-Xmx11674m	-Xmx11674m
mapreduce.map.memory.mb	7296	7296
mapreduce.reduce.memory.mb	14592	14592
yarn.app.mapreduce.am.resource.mb	14592	14592
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## i4i.8xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12084m</code>	<code>-Xmx12084m</code>
<code>mapreduce.map.memory.mb</code>	7552	7552
<code>mapreduce.reduce.memory.mb</code>	15104	15104
<code>yarn.app.mapreduce.am.resource.mb</code>	15104	15104
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

## i4i.12xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6110m</code>	<code>-Xmx6110m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12220m</code>	<code>-Xmx12220m</code>
<code>mapreduce.map.memory.mb</code>	7637	7637
<code>mapreduce.reduce.memory.mb</code>	15274	15274
<code>yarn.app.mapreduce.am.resource.mb</code>	15274	15274



Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	366592	30564
<code>yarn.nodemanager.resource.memory-mb</code>	366592	30564

## i4i.16xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6144m</code>	<code>-Xmx6144m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12288m</code>	<code>-Xmx12288m</code>
<code>mapreduce.map.memory.mb</code>	7680	7680
<code>mapreduce.reduce.memory.mb</code>	15360	15360
<code>yarn.app.mapreduce.am.resource.mb</code>	15360	15360
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	491520	30720
<code>yarn.nodemanager.resource.memory-mb</code>	491520	30720

## i4i.24xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6178m</code>	<code>-Xmx6178m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12356m</code>	<code>-Xmx12356m</code>
<code>mapreduce.map.memory.mb</code>	7723	7723
<code>mapreduce.reduce.memory.mb</code>	15446	15446
<code>yarn.app.mapreduce.am.resource.mb</code>	15446	15446
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	741376	30860
<code>yarn.nodemanager.resource.memory-mb</code>	741376	30860

## i4i.32xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6195m</code>	<code>-Xmx6195m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12390m</code>	<code>-Xmx12390m</code>
<code>mapreduce.map.memory.mb</code>	7744	7744
<code>mapreduce.reduce.memory.mb</code>	15488	15488
<code>yarn.app.mapreduce.am.resource.mb</code>	15488	15488

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	991232	30976
<code>yarn.nodemanager.resource.memory-mb</code>	991232	30976

## Im4gn-Instances

### im4gn.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4684m</code>	<code>-Xmx4684m</code>
<code>mapreduce.map.memory.mb</code>	2928	2928
<code>mapreduce.reduce.memory.mb</code>	5856	5856
<code>yarn.app.mapreduce.am.resource.mb</code>	5856	5856
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	11712	5856
<code>yarn.nodemanager.resource.memory-mb</code>	11712	5856

## im4gn.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## im4gn.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.java.opts	-Xmx5428m	-Xmx5428m
mapreduce.map.memory.mb	3392	3392
mapreduce.reduce.memory.mb	6784	6784
yarn.app.mapreduce.am.resource.mb	6784	6784

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

## im4gn.8xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx2918m</code>	<code>-Xmx2918m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5836m</code>	<code>-Xmx5836m</code>
<code>mapreduce.map.memory.mb</code>	3648	3648
<code>mapreduce.reduce.memory.mb</code>	7296	7296
<code>yarn.app.mapreduce.am.resource.mb</code>	7296	7296
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

## im4gn.16xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx3021m	-Xmx3021m
mapreduce.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.map.memory.mb	3776	3776
mapreduce.reduce.memory.mb	7552	7552
yarn.app.mapreduce.am.resource.mb	7552	7552
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

## Is4gen-Instances

## is4gen.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx3514m	-Xmx3514m
mapreduce.java.opts	-Xmx7028m	-Xmx7028m
mapreduce.map.memory.mb	4393	4393
mapreduce.reduce.memory.mb	8786	8786

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.app.mapreduce.am.resource.mb	8786	8786
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	17572,12	8786,06
yarn.nodemanager.resource.memory-mb	17572,12	8786,06

## is4gen.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx3866m	-Xmx3866m
mapreduce.java.opts	-Xmx7732m	-Xmx7732m
mapreduce.map.memory.mb	4832	4832
mapreduce.reduce.memory.mb	9664	9664
yarn.app.mapreduce.am.resource.mb	9664	9664
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	38656	19328
yarn.nodemanager.resource.memory-mb	38656	19328

## is4gen.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx4275m	-Xmx4275m
mapreduce.java.opts	-Xmx8550m	-Xmx8550m
mapreduce.map.memory.mb	5344	5344
mapreduce.reduce.memory.mb	10688	10688
yarn.app.mapreduce.am.resource.mb	10688	10688
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	85504	32064
yarn.nodemanager.resource.memory-mb	85504	32064

## is4gen.8xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx4480m	-Xmx4480m
mapreduce.java.opts	-Xmx8960m	-Xmx8960m
mapreduce.map.memory.mb	5600	5600
mapreduce.reduce.memory.mb	11200	11200
yarn.app.mapreduce.am.resource.mb	11200	11200



Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	179200	22400
<code>yarn.nodemanager.resource.memory-mb</code>	179200	22400

## M1-Instances

### m1.small

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx288m</code>	<code>-Xmx288m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx288m</code>	<code>-Xmx288m</code>
<code>mapreduce.map.memory.mb</code>	512	512
<code>mapreduce.reduce.memory.mb</code>	512	512
<code>yarn.app.mapreduce.am.resource.mb</code>	512	512
<code>yarn.scheduler.minimum-allocation-mb</code>	256	256
<code>yarn.scheduler.maximum-allocation-mb</code>	512	512
<code>yarn.nodemanager.resource.memory-mb</code>	1024	512

## m1.medium

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx512m</code>	<code>-Xmx512m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx768m</code>	<code>-Xmx768m</code>
<code>mapreduce.map.memory.mb</code>	768	768
<code>mapreduce.reduce.memory.mb</code>	1024	1024
<code>yarn.app.mapreduce.am.resource.mb</code>	1024	1024
<code>yarn.scheduler.minimum-allocation-mb</code>	256	256
<code>yarn.scheduler.maximum-allocation-mb</code>	2048	1024
<code>yarn.nodemanager.resource.memory-mb</code>	2048	1024

## m1.large

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx512m</code>	<code>-Xmx512m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx1024m</code>	<code>-Xmx1024m</code>
<code>mapreduce.map.memory.mb</code>	768	768
<code>mapreduce.reduce.memory.mb</code>	1536	1536
<code>yarn.app.mapreduce.am.resource.mb</code>	1536	1536

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	256	256
<code>yarn.scheduler.maximum-allocation-mb</code>	3072	2560
<code>yarn.nodemanager.resource.memory-mb</code>	5120	2560

## m1.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx512m</code>	<code>-Xmx512m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx1536m</code>	<code>-Xmx1536m</code>
<code>mapreduce.map.memory.mb</code>	768	768
<code>mapreduce.reduce.memory.mb</code>	2048	2048
<code>yarn.app.mapreduce.am.resource.mb</code>	2048	2048
<code>yarn.scheduler.minimum-allocation-mb</code>	256	256
<code>yarn.scheduler.maximum-allocation-mb</code>	8192	6144
<code>yarn.nodemanager.resource.memory-mb</code>	12288	6144

## M2-Instances

## m2.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx864m	-Xmx864m
mapreduce.java.opts	-Xmx1536m	-Xmx1536m
mapreduce.map.memory.mb	1024	1024
mapreduce.reduce.memory.mb	2048	2048
yarn.app.mapreduce.am.resource.mb	2048	2048
yarn.scheduler.minimum-allocation-mb	256	256
yarn.scheduler.maximum-allocation-mb	7168	7168
yarn.nodemanager.resource.memory-mb	14336	7168

## m2.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1280m	-Xmx1280m
mapreduce.java.opts	-Xmx2304m	-Xmx2304m
mapreduce.map.memory.mb	1536	1536
mapreduce.reduce.memory.mb	2560	2560

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.app.mapreduce.am.resource.mb</code>	2560	2560
<code>yarn.scheduler.minimum-allocation-mb</code>	256	256
<code>yarn.scheduler.maximum-allocation-mb</code>	8192	15360
<code>yarn.nodemanager.resource.memory-mb</code>	30720	15360

## m2.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx1280m</code>	<code>-Xmx1280m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2304m</code>	<code>-Xmx2304m</code>
<code>mapreduce.map.memory.mb</code>	1536	1536
<code>mapreduce.reduce.memory.mb</code>	2560	2560
<code>yarn.app.mapreduce.am.resource.mb</code>	2560	2560
<code>yarn.scheduler.minimum-allocation-mb</code>	256	256
<code>yarn.scheduler.maximum-allocation-mb</code>	8192	30720
<code>yarn.nodemanager.resource.memory-mb</code>	61440	30720

## M3-Instances

## m3.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1152m	-Xmx1152m
mapreduce.java.opts	-Xmx2304m	-Xmx2304m
mapreduce.map.memory.mb	1440	1440
mapreduce.reduce.memory.mb	2880	2880
yarn.app.mapreduce.am.resource.mb	2880	2880
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11520	5760
yarn.nodemanager.resource.memory-mb	11520	5760

## m3.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1152m	-Xmx1152m
mapreduce.java.opts	-Xmx2304m	-Xmx2304m
mapreduce.map.memory.mb	1440	1440
mapreduce.reduce.memory.mb	2880	2880

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.app.mapreduce.am.resource.mb	2880	2880
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23040	11520
yarn.nodemanager.resource.memory-mb	23040	11520

## M4-Instances

### m4.large

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1229m	-Xmx1229m
mapreduce.java.opts	-Xmx2458m	-Xmx2458m
mapreduce.map.memory.mb	1536	1536
mapreduce.reduce.memory.mb	3072	3072
yarn.app.mapreduce.am.resource.mb	3072	3072
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	6144	3072

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.nodemanager.resource.memory-mb</code>	6144	3072

## m4.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx1229m</code>	<code>-Xmx1229m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2458m</code>	<code>-Xmx2458m</code>
<code>mapreduce.map.memory.mb</code>	1536	1536
<code>mapreduce.reduce.memory.mb</code>	3072	3072
<code>yarn.app.mapreduce.am.resource.mb</code>	3072	3072
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	12288	6144
<code>yarn.nodemanager.resource.memory-mb</code>	12288	6144

## m4.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx1229m</code>	<code>-Xmx1229m</code>



Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.java.opts</code>	<code>-Xmx2458m</code>	<code>-Xmx2458m</code>
<code>mapreduce.map.memory.mb</code>	1536	1536
<code>mapreduce.reduce.memory.mb</code>	3072	3072
<code>yarn.app.mapreduce.am.resource.mb</code>	3072	3072
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	24576	12288
<code>yarn.nodemanager.resource.memory-mb</code>	24576	12288

#### m4.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx1434m</code>	<code>-Xmx1434m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2868m</code>	<code>-Xmx2868m</code>
<code>mapreduce.map.memory.mb</code>	1792	1792
<code>mapreduce.reduce.memory.mb</code>	3584	3584
<code>yarn.app.mapreduce.am.resource.mb</code>	3584	3584
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.maximum-allocation-mb</code>	57344	28672
<code>yarn.nodemanager.resource.memory-mb</code>	57344	28672

## m4.10xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx1557m</code>	<code>-Xmx1557m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx3114m</code>	<code>-Xmx3114m</code>
<code>mapreduce.map.memory.mb</code>	1946	1946
<code>mapreduce.reduce.memory.mb</code>	3892	3892
<code>yarn.app.mapreduce.am.resource.mb</code>	3892	3892
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	155648	31104
<code>yarn.nodemanager.resource.memory-mb</code>	155648	31104

## m4.16xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx1587m	-Xmx1587m
mapreduce.java.opts	-Xmx3174m	-Xmx3174m
mapreduce.map.memory.mb	1984	1984
mapreduce.reduce.memory.mb	3968	3968
yarn.app.mapreduce.am.resource.mb	3968	3968
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	253952	31744
yarn.nodemanager.resource.memory-mb	253952	31744

## M5-Instances

## m5.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx2458m	-Xmx2458m
mapreduce.java.opts	-Xmx4916m	-Xmx4916m
mapreduce.map.memory.mb	3072	3072
mapreduce.reduce.memory.mb	6144	6144

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.app.mapreduce.am.resource.mb	6144	6144
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	12288	6144
yarn.nodemanager.resource.memory-mb	12288	6144

## m5.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx2458m	-Xmx2458m
mapreduce.java.opts	-Xmx4916m	-Xmx4916m
mapreduce.map.memory.mb	3072	3072
mapreduce.reduce.memory.mb	6144	6144
yarn.app.mapreduce.am.resource.mb	6144	6144
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	24576	12288
yarn.nodemanager.resource.memory-mb	24576	12288

## m5.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx2867m</code>	<code>-Xmx2867m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5734m</code>	<code>-Xmx5734m</code>
<code>mapreduce.map.memory.mb</code>	3584	3584
<code>mapreduce.reduce.memory.mb</code>	7168	7168
<code>yarn.app.mapreduce.am.resource.mb</code>	7168	7168
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	57344	28672
<code>yarn.nodemanager.resource.memory-mb</code>	57344	28672

## m5.8xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx3072m</code>	<code>-Xmx3072m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6144m</code>	<code>-Xmx6144m</code>
<code>mapreduce.map.memory.mb</code>	3840	3840
<code>mapreduce.reduce.memory.mb</code>	7680	7680
<code>yarn.app.mapreduce.am.resource.mb</code>	7680	7680

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	122880	30720
<code>yarn.nodemanager.resource.memory-mb</code>	122880	30720

## m5.12xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx3140m</code>	<code>-Xmx3140m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6280m</code>	<code>-Xmx6280m</code>
<code>mapreduce.map.memory.mb</code>	3925	3925
<code>mapreduce.reduce.memory.mb</code>	7850	7850
<code>yarn.app.mapreduce.am.resource.mb</code>	7850	7850
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	188416	31416
<code>yarn.nodemanager.resource.memory-mb</code>	188416	31416

## m5.16xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx3174m	-Xmx3174m
mapreduce.java.opts	-Xmx6348m	-Xmx6348m
mapreduce.map.memory.mb	3968	3968
mapreduce.reduce.memory.mb	7936	7936
yarn.app.mapreduce.am.resource.mb	7936	7936
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	253952	31744
yarn.nodemanager.resource.memory-mb	253952	31744

## m5.24xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx3209m	-Xmx3209m
mapreduce.java.opts	-Xmx6418m	-Xmx6418m
mapreduce.map.memory.mb	4011	4011
mapreduce.reduce.memory.mb	8022	8022
yarn.app.mapreduce.am.resource.mb	8022	8022

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	385024	32056
<code>yarn.nodemanager.resource.memory-mb</code>	385024	32056

## M5a-Instances

### m5a.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx2458m</code>	<code>-Xmx2458m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4916m</code>	<code>-Xmx4916m</code>
<code>mapreduce.map.memory.mb</code>	3072	3072
<code>mapreduce.reduce.memory.mb</code>	6144	6144
<code>yarn.app.mapreduce.am.resource.mb</code>	6144	6144
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	12288	6144
<code>yarn.nodemanager.resource.memory-mb</code>	12288	6144



## m5a.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx2458m</code>	<code>-Xmx2458m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4916m</code>	<code>-Xmx4916m</code>
<code>mapreduce.map.memory.mb</code>	3072	3072
<code>mapreduce.reduce.memory.mb</code>	6144	6144
<code>yarn.app.mapreduce.am.resource.mb</code>	6144	6144
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	24576	12288
<code>yarn.nodemanager.resource.memory-mb</code>	24576	12288

## m5a.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx2867m</code>	<code>-Xmx2867m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5734m</code>	<code>-Xmx5734m</code>
<code>mapreduce.map.memory.mb</code>	3584	3584
<code>mapreduce.reduce.memory.mb</code>	7168	7168
<code>yarn.app.mapreduce.am.resource.mb</code>	7168	7168

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	57344	28672
<code>yarn.nodemanager.resource.memory-mb</code>	57344	28672

## m5a.8xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx3072m</code>	<code>-Xmx3072m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6144m</code>	<code>-Xmx6144m</code>
<code>mapreduce.map.memory.mb</code>	3840	3840
<code>mapreduce.reduce.memory.mb</code>	7680	7680
<code>yarn.app.mapreduce.am.resource.mb</code>	7680	7680
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	122880	30720
<code>yarn.nodemanager.resource.memory-mb</code>	122880	30720

## m5a.12xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx3140m</code>	<code>-Xmx3140m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6280m</code>	<code>-Xmx6280m</code>
<code>mapreduce.map.memory.mb</code>	3925	3925
<code>mapreduce.reduce.memory.mb</code>	7850	7850
<code>yarn.app.mapreduce.am.resource.mb</code>	7850	7850
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	188416	31416
<code>yarn.nodemanager.resource.memory-mb</code>	188416	31416

## m5a.16xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx3174m</code>	<code>-Xmx3174m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6348m</code>	<code>-Xmx6348m</code>
<code>mapreduce.map.memory.mb</code>	3968	3968
<code>mapreduce.reduce.memory.mb</code>	7936	7936
<code>yarn.app.mapreduce.am.resource.mb</code>	7936	7936

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	253952	31744
<code>yarn.nodemanager.resource.memory-mb</code>	253952	31744

## m5a.24xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx3209m</code>	<code>-Xmx3209m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6418m</code>	<code>-Xmx6418m</code>
<code>mapreduce.map.memory.mb</code>	4011	4011
<code>mapreduce.reduce.memory.mb</code>	8022	8022
<code>yarn.app.mapreduce.am.resource.mb</code>	8022	8022
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	385024	32056
<code>yarn.nodemanager.resource.memory-mb</code>	385024	32056

## M5ad-Instances

## m5ad.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

## m5ad.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## m5ad.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.java.opts	-Xmx5428m	-Xmx5428m
mapreduce.map.memory.mb	3392	3392
mapreduce.reduce.memory.mb	6784	6784
yarn.app.mapreduce.am.resource.mb	6784	6784
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## m5ad.8xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.java.opts	-Xmx5836m	-Xmx5836m
mapreduce.map.memory.mb	3648	3648
mapreduce.reduce.memory.mb	7296	7296
yarn.app.mapreduce.am.resource.mb	7296	7296
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## m5ad.12xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx2986m	-Xmx2986m
mapreduce.java.opts	-Xmx5972m	-Xmx5972m
mapreduce.map.memory.mb	3733	3733
mapreduce.reduce.memory.mb	7466	7466
yarn.app.mapreduce.am.resource.mb	7466	7466

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	179200	29880
<code>yarn.nodemanager.resource.memory-mb</code>	179200	29880

## m5ad.16xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx3021m</code>	<code>-Xmx3021m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.map.memory.mb</code>	3776	3776
<code>mapreduce.reduce.memory.mb</code>	7552	7552
<code>yarn.app.mapreduce.am.resource.mb</code>	7552	7552
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208



## m5ad.24xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx3055m	-Xmx3055m
mapreduce.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.map.memory.mb	3819	3819
mapreduce.reduce.memory.mb	7638	7638
yarn.app.mapreduce.am.resource.mb	7638	7638
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30520
yarn.nodemanager.resource.memory-mb	366592	30520

## M5d-Instances

## m5d.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx2458m	-Xmx2458m
mapreduce.java.opts	-Xmx4916m	-Xmx4916m
mapreduce.map.memory.mb	3072	3072
mapreduce.reduce.memory.mb	6144	6144

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.app.mapreduce.am.resource.mb	6144	6144
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	12288	6144
yarn.nodemanager.resource.memory-mb	12288	6144

## m5d.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx2458m	-Xmx2458m
mapreduce.java.opts	-Xmx4916m	-Xmx4916m
mapreduce.map.memory.mb	3072	3072
mapreduce.reduce.memory.mb	6144	6144
yarn.app.mapreduce.am.resource.mb	6144	6144
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	24576	12288
yarn.nodemanager.resource.memory-mb	24576	12288

## m5d.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx2867m</code>	<code>-Xmx2867m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5734m</code>	<code>-Xmx5734m</code>
<code>mapreduce.map.memory.mb</code>	3584	3584
<code>mapreduce.reduce.memory.mb</code>	7168	7168
<code>yarn.app.mapreduce.am.resource.mb</code>	7168	7168
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	57344	28672
<code>yarn.nodemanager.resource.memory-mb</code>	57344	28672

## m5d.8xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx3072m</code>	<code>-Xmx3072m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6144m</code>	<code>-Xmx6144m</code>
<code>mapreduce.map.memory.mb</code>	3840	3840
<code>mapreduce.reduce.memory.mb</code>	7680	7680
<code>yarn.app.mapreduce.am.resource.mb</code>	7680	7680

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	122880	30720
<code>yarn.nodemanager.resource.memory-mb</code>	122880	30720

## m5d.12xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx3140m</code>	<code>-Xmx3140m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6280m</code>	<code>-Xmx6280m</code>
<code>mapreduce.map.memory.mb</code>	3925	3925
<code>mapreduce.reduce.memory.mb</code>	7850	7850
<code>yarn.app.mapreduce.am.resource.mb</code>	7850	7850
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	188416	31416
<code>yarn.nodemanager.resource.memory-mb</code>	188416	31416

## m5d.16xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx3174m	-Xmx3174m
mapreduce.java.opts	-Xmx6348m	-Xmx6348m
mapreduce.map.memory.mb	3968	3968
mapreduce.reduce.memory.mb	7936	7936
yarn.app.mapreduce.am.resource.mb	7936	7936
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	253952	31744
yarn.nodemanager.resource.memory-mb	253952	31744

## m5d.24xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx3209m	-Xmx3209m
mapreduce.java.opts	-Xmx6418m	-Xmx6418m
mapreduce.map.memory.mb	4011	4011
mapreduce.reduce.memory.mb	8022	8022
yarn.app.mapreduce.am.resource.mb	8022	8022

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	385024	32056
<code>yarn.nodemanager.resource.memory-mb</code>	385024	32056

## M5dn-Instances

### m5dn.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4684m</code>	<code>-Xmx4684m</code>
<code>mapreduce.map.memory.mb</code>	2928	2928
<code>mapreduce.reduce.memory.mb</code>	5856	5856
<code>yarn.app.mapreduce.am.resource.mb</code>	5856	5856
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	11712	5856
<code>yarn.nodemanager.resource.memory-mb</code>	11712	5856

## m5dn.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## m5dn.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.java.opts	-Xmx5428m	-Xmx5428m
mapreduce.map.memory.mb	3392	3392
mapreduce.reduce.memory.mb	6784	6784
yarn.app.mapreduce.am.resource.mb	6784	6784

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

## m5dn.8xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx2918m</code>	<code>-Xmx2918m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5836m</code>	<code>-Xmx5836m</code>
<code>mapreduce.map.memory.mb</code>	3648	3648
<code>mapreduce.reduce.memory.mb</code>	7296	7296
<code>yarn.app.mapreduce.am.resource.mb</code>	7296	7296
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184



## m5dn.12xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx2986m	-Xmx2986m
mapreduce.java.opts	-Xmx5972m	-Xmx5972m
mapreduce.map.memory.mb	3733	3733
mapreduce.reduce.memory.mb	7466	7466
yarn.app.mapreduce.am.resource.mb	7466	7466
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	179200	29880
yarn.nodemanager.resource.memory-mb	179200	29880

## m5dn.16xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx3021m	-Xmx3021m
mapreduce.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.map.memory.mb	3776	3776
mapreduce.reduce.memory.mb	7552	7552
yarn.app.mapreduce.am.resource.mb	7552	7552

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

## m5dn.24xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx3055m</code>	<code>-Xmx3055m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6110m</code>	<code>-Xmx6110m</code>
<code>mapreduce.map.memory.mb</code>	3819	3819
<code>mapreduce.reduce.memory.mb</code>	7638	7638
<code>yarn.app.mapreduce.am.resource.mb</code>	7638	7638
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	366592	30520
<code>yarn.nodemanager.resource.memory-mb</code>	366592	30520

## M5n-Instances

## m5n.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

## m5n.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## m5n.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.java.opts	-Xmx5428m	-Xmx5428m
mapreduce.map.memory.mb	3392	3392
mapreduce.reduce.memory.mb	6784	6784
yarn.app.mapreduce.am.resource.mb	6784	6784
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## m5n.8xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.java.opts	-Xmx5836m	-Xmx5836m
mapreduce.map.memory.mb	3648	3648
mapreduce.reduce.memory.mb	7296	7296
yarn.app.mapreduce.am.resource.mb	7296	7296
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## m5n.12xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx2986m	-Xmx2986m
mapreduce.java.opts	-Xmx5972m	-Xmx5972m
mapreduce.map.memory.mb	3733	3733
mapreduce.reduce.memory.mb	7466	7466
yarn.app.mapreduce.am.resource.mb	7466	7466

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	179200	29880
<code>yarn.nodemanager.resource.memory-mb</code>	179200	29880

## m5n.16xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx3021m</code>	<code>-Xmx3021m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.map.memory.mb</code>	3776	3776
<code>mapreduce.reduce.memory.mb</code>	7552	7552
<code>yarn.app.mapreduce.am.resource.mb</code>	7552	7552
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

## m5n.24xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx3055m	-Xmx3055m
mapreduce.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.map.memory.mb	3819	3819
mapreduce.reduce.memory.mb	7638	7638
yarn.app.mapreduce.am.resource.mb	7638	7638
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30520
yarn.nodemanager.resource.memory-mb	366592	30520

## M5zn-Instances

## m5zn.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx2304m	-Xmx2304m
mapreduce.java.opts	-Xmx4608m	-Xmx4608m
mapreduce.map.memory.mb	2880	2880
mapreduce.reduce.memory.mb	5760	5760

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.app.mapreduce.am.resource.mb	5760	5760
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11520	5760
yarn.nodemanager.resource.memory-mb	11520	5760

## m5zn.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712



## m5zn.3xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx2577m	-Xmx2577m
mapreduce.java.opts	-Xmx5154m	-Xmx5154m
mapreduce.map.memory.mb	3221	3221
mapreduce.reduce.memory.mb	6442	6442
yarn.app.mapreduce.am.resource.mb	6442	6442
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	38656	19328
yarn.nodemanager.resource.memory-mb	38656	19328

## m5zn.6xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx2850m	-Xmx2850m
mapreduce.java.opts	-Xmx5700m	-Xmx5700m
mapreduce.map.memory.mb	3563	3563
mapreduce.reduce.memory.mb	7126	7126
yarn.app.mapreduce.am.resource.mb	7126	7126

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	85504	28496
<code>yarn.nodemanager.resource.memory-mb</code>	85504	28496

## m5zn.12xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx2986m</code>	<code>-Xmx2986m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5972m</code>	<code>-Xmx5972m</code>
<code>mapreduce.map.memory.mb</code>	3733	3733
<code>mapreduce.reduce.memory.mb</code>	7466	7466
<code>yarn.app.mapreduce.am.resource.mb</code>	7466	7466
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	179200	29880
<code>yarn.nodemanager.resource.memory-mb</code>	179200	29880

## M6a-Instances

## m6a.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

## m6a.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## m6a.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.java.opts	-Xmx5428m	-Xmx5428m
mapreduce.map.memory.mb	3392	3392
mapreduce.reduce.memory.mb	6784	6784
yarn.app.mapreduce.am.resource.mb	6784	6784
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## m6a.8xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.java.opts	-Xmx5836m	-Xmx5836m
mapreduce.map.memory.mb	3648	3648
mapreduce.reduce.memory.mb	7296	7296
yarn.app.mapreduce.am.resource.mb	7296	7296
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## m6a.12xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx2986m	-Xmx2986m
mapreduce.java.opts	-Xmx5972m	-Xmx5972m
mapreduce.map.memory.mb	3733	3733
mapreduce.reduce.memory.mb	7466	7466
yarn.app.mapreduce.am.resource.mb	7466	7466

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	179200	29880
<code>yarn.nodemanager.resource.memory-mb</code>	179200	29880

## m6a.16xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx3021m</code>	<code>-Xmx3021m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.map.memory.mb</code>	3776	3776
<code>mapreduce.reduce.memory.mb</code>	7552	7552
<code>yarn.app.mapreduce.am.resource.mb</code>	7552	7552
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

## m6a.24xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx3055m	-Xmx3055m
mapreduce.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.map.memory.mb	3819	3819
mapreduce.reduce.memory.mb	7638	7638
yarn.app.mapreduce.am.resource.mb	7638	7638
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30520
yarn.nodemanager.resource.memory-mb	366592	30520

## m6a.32xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx3072m	-Xmx3072m
mapreduce.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.map.memory.mb	3840	3840
mapreduce.reduce.memory.mb	7680	7680
yarn.app.mapreduce.am.resource.mb	7680	7680

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	491520	30720
<code>yarn.nodemanager.resource.memory-mb</code>	491520	30720

## m6a.48xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx3089m</code>	<code>-Xmx3089m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6178m</code>	<code>-Xmx6178m</code>
<code>mapreduce.map.memory.mb</code>	3861	3861
<code>mapreduce.reduce.memory.mb</code>	7722	7722
<code>yarn.app.mapreduce.am.resource.mb</code>	7722	7722
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	741376	30952
<code>yarn.nodemanager.resource.memory-mb</code>	741376	30952



## M6g-Instances

## m6g.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

## m6g.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## m6g.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.java.opts	-Xmx5428m	-Xmx5428m
mapreduce.map.memory.mb	3392	3392
mapreduce.reduce.memory.mb	6784	6784
yarn.app.mapreduce.am.resource.mb	6784	6784
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## m6g.8xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.java.opts	-Xmx5836m	-Xmx5836m
mapreduce.map.memory.mb	3648	3648
mapreduce.reduce.memory.mb	7296	7296
yarn.app.mapreduce.am.resource.mb	7296	7296
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## m6g.12xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx3021m	-Xmx3021m
mapreduce.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.map.memory.mb	3776	3776
mapreduce.reduce.memory.mb	7552	7552
yarn.app.mapreduce.am.resource.mb	7552	7552

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	181248	30208
<code>yarn.nodemanager.resource.memory-mb</code>	181248	30208

## m6g.16xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx3021m</code>	<code>-Xmx3021m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.map.memory.mb</code>	3776	3776
<code>mapreduce.reduce.memory.mb</code>	7552	7552
<code>yarn.app.mapreduce.am.resource.mb</code>	7552	7552
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

## M6gd-Instances

## m6gd.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

## m6gd.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## m6gd.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.java.opts	-Xmx5428m	-Xmx5428m
mapreduce.map.memory.mb	3392	3392
mapreduce.reduce.memory.mb	6784	6784
yarn.app.mapreduce.am.resource.mb	6784	6784
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## m6gd.8xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.java.opts	-Xmx5836m	-Xmx5836m
mapreduce.map.memory.mb	3648	3648
mapreduce.reduce.memory.mb	7296	7296
yarn.app.mapreduce.am.resource.mb	7296	7296
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## m6gd.12xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx3021m	-Xmx3021m
mapreduce.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.map.memory.mb	3776	3776
mapreduce.reduce.memory.mb	7552	7552
yarn.app.mapreduce.am.resource.mb	7552	7552

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	181248	30208
<code>yarn.nodemanager.resource.memory-mb</code>	181248	30208

## m6gd.16xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx3021m</code>	<code>-Xmx3021m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.map.memory.mb</code>	3776	3776
<code>mapreduce.reduce.memory.mb</code>	7552	7552
<code>yarn.app.mapreduce.am.resource.mb</code>	7552	7552
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208



## M6i-Instances

## m6i.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

## m6i.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## m6i.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.java.opts	-Xmx5428m	-Xmx5428m
mapreduce.map.memory.mb	3392	3392
mapreduce.reduce.memory.mb	6784	6784
yarn.app.mapreduce.am.resource.mb	6784	6784
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## m6i.8xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx2918m</code>	<code>-Xmx2918m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5836m</code>	<code>-Xmx5836m</code>
<code>mapreduce.map.memory.mb</code>	3648	3648
<code>mapreduce.reduce.memory.mb</code>	7296	7296
<code>yarn.app.mapreduce.am.resource.mb</code>	7296	7296
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

## m6i.12xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx3021m</code>	<code>-Xmx3021m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.map.memory.mb</code>	3776	3776
<code>mapreduce.reduce.memory.mb</code>	7552	7552
<code>yarn.app.mapreduce.am.resource.mb</code>	7552	7552

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	181248	30208
<code>yarn.nodemanager.resource.memory-mb</code>	181248	30208

## m6i.16xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx3021m</code>	<code>-Xmx3021m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.map.memory.mb</code>	3776	3776
<code>mapreduce.reduce.memory.mb</code>	7552	7552
<code>yarn.app.mapreduce.am.resource.mb</code>	7552	7552
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

## m6i.24xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx3055m</code>	<code>-Xmx3055m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6110m</code>	<code>-Xmx6110m</code>
<code>mapreduce.map.memory.mb</code>	3819	3819
<code>mapreduce.reduce.memory.mb</code>	7638	7638
<code>yarn.app.mapreduce.am.resource.mb</code>	7638	7638
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	366592	30520
<code>yarn.nodemanager.resource.memory-mb</code>	366592	30520

## m6i.32xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx3072m</code>	<code>-Xmx3072m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6144m</code>	<code>-Xmx6144m</code>
<code>mapreduce.map.memory.mb</code>	3840	3840
<code>mapreduce.reduce.memory.mb</code>	7680	7680
<code>yarn.app.mapreduce.am.resource.mb</code>	7680	7680

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	491520	30720
<code>yarn.nodemanager.resource.memory-mb</code>	491520	30720

## M6id-Instances

### m6id.x groß

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4684m</code>	<code>-Xmx4684m</code>
<code>mapreduce.map.memory.mb</code>	2928	2928
<code>mapreduce.reduce.memory.mb</code>	5856	5856
<code>yarn.app.mapreduce.am.resource.mb</code>	5856	5856
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	11712	5856
<code>yarn.nodemanager.resource.memory-mb</code>	11712	5856

## m6id.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4684m</code>	<code>-Xmx4684m</code>
<code>mapreduce.map.memory.mb</code>	2928	2928
<code>mapreduce.reduce.memory.mb</code>	5856	5856
<code>yarn.app.mapreduce.am.resource.mb</code>	5856	5856
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

## m6id.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx2714m</code>	<code>-Xmx2714m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5428m</code>	<code>-Xmx5428m</code>
<code>mapreduce.map.memory.mb</code>	3392	3392
<code>mapreduce.reduce.memory.mb</code>	6784	6784
<code>yarn.app.mapreduce.am.resource.mb</code>	6784	6784

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

## m6id.8xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx2918m</code>	<code>-Xmx2918m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5836m</code>	<code>-Xmx5836m</code>
<code>mapreduce.map.memory.mb</code>	3648	3648
<code>mapreduce.reduce.memory.mb</code>	7296	7296
<code>yarn.app.mapreduce.am.resource.mb</code>	7296	7296
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184



## m6id.12xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx2986m</code>	<code>-Xmx2986m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5972m</code>	<code>-Xmx5972m</code>
<code>mapreduce.map.memory.mb</code>	3733	3733
<code>mapreduce.reduce.memory.mb</code>	7466	7466
<code>yarn.app.mapreduce.am.resource.mb</code>	7466	7466
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	179200	29880
<code>yarn.nodemanager.resource.memory-mb</code>	179200	29880

## m6id.16xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx3021m</code>	<code>-Xmx3021m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.map.memory.mb</code>	3776	3776
<code>mapreduce.reduce.memory.mb</code>	7552	7552
<code>yarn.app.mapreduce.am.resource.mb</code>	7552	7552

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

## m6id.24xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx3055m</code>	<code>-Xmx3055m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6110m</code>	<code>-Xmx6110m</code>
<code>mapreduce.map.memory.mb</code>	3819	3819
<code>mapreduce.reduce.memory.mb</code>	7638	7638
<code>yarn.app.mapreduce.am.resource.mb</code>	7638	7638
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	366592	30520
<code>yarn.nodemanager.resource.memory-mb</code>	366592	30520

## m6id.32xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx3072m	-Xmx3072m
mapreduce.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.map.memory.mb	3840	3840
mapreduce.reduce.memory.mb	7680	7680
yarn.app.mapreduce.am.resource.mb	7680	7680
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

## M6idn-Instances

## m6idn.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

## m6id.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## m6idn.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx2714m</code>	<code>-Xmx2714m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5428m</code>	<code>-Xmx5428m</code>
<code>mapreduce.map.memory.mb</code>	3392	3392
<code>mapreduce.reduce.memory.mb</code>	6784	6784
<code>yarn.app.mapreduce.am.resource.mb</code>	6784	6784
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

## m6idn.8xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx2918m</code>	<code>-Xmx2918m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5836m</code>	<code>-Xmx5836m</code>
<code>mapreduce.map.memory.mb</code>	3648	3648
<code>mapreduce.reduce.memory.mb</code>	7296	7296
<code>yarn.app.mapreduce.am.resource.mb</code>	7296	7296

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

## m6idn.12xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx2986m</code>	<code>-Xmx2986m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5972m</code>	<code>-Xmx5972m</code>
<code>mapreduce.map.memory.mb</code>	3733	3733
<code>mapreduce.reduce.memory.mb</code>	7466	7466
<code>yarn.app.mapreduce.am.resource.mb</code>	7466	7466
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	179200	29880
<code>yarn.nodemanager.resource.memory-mb</code>	179200	29880

## m6idn.16xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx3021m	-Xmx3021m
mapreduce.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.map.memory.mb	3776	3776
mapreduce.reduce.memory.mb	7552	7552
yarn.app.mapreduce.am.resource.mb	7552	7552
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

## m6idn.24xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx3055m	-Xmx3055m
mapreduce.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.map.memory.mb	3819	3819
mapreduce.reduce.memory.mb	7638	7638
yarn.app.mapreduce.am.resource.mb	7638	7638

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	366592	30520
<code>yarn.nodemanager.resource.memory-mb</code>	366592	30520

## m6idn.32xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx3072m</code>	<code>-Xmx3072m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6144m</code>	<code>-Xmx6144m</code>
<code>mapreduce.map.memory.mb</code>	3840	3840
<code>mapreduce.reduce.memory.mb</code>	7680	7680
<code>yarn.app.mapreduce.am.resource.mb</code>	7680	7680
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	491520	30720
<code>yarn.nodemanager.resource.memory-mb</code>	491520	30720



## M6in-Instances

## m6in.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

## m6in.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## m6in.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.java.opts	-Xmx5428m	-Xmx5428m
mapreduce.map.memory.mb	3392	3392
mapreduce.reduce.memory.mb	6784	6784
yarn.app.mapreduce.am.resource.mb	6784	6784
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## m6n.8xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.java.opts	-Xmx5836m	-Xmx5836m
mapreduce.map.memory.mb	3648	3648
mapreduce.reduce.memory.mb	7296	7296
yarn.app.mapreduce.am.resource.mb	7296	7296
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## m6in.12xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx3021m	-Xmx3021m
mapreduce.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.map.memory.mb	3776	3776
mapreduce.reduce.memory.mb	7552	7552
yarn.app.mapreduce.am.resource.mb	7552	7552

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	181248	30208
<code>yarn.nodemanager.resource.memory-mb</code>	181248	30208

## m6in.16xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx3021m</code>	<code>-Xmx3021m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.map.memory.mb</code>	3776	3776
<code>mapreduce.reduce.memory.mb</code>	7552	7552
<code>yarn.app.mapreduce.am.resource.mb</code>	7552	7552
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

## m6in.24xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx3055m	-Xmx3055m
mapreduce.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.map.memory.mb	3819	3819
mapreduce.reduce.memory.mb	7638	7638
yarn.app.mapreduce.am.resource.mb	7638	7638
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30520
yarn.nodemanager.resource.memory-mb	366592	30520

## m6in.32xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx3072m	-Xmx3072m
mapreduce.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.map.memory.mb	3840	3840
mapreduce.reduce.memory.mb	7680	7680
yarn.app.mapreduce.am.resource.mb	7680	7680

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	491520	30720
<code>yarn.nodemanager.resource.memory-mb</code>	491520	30720

## M7a-Instances

### m7a.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4684m</code>	<code>-Xmx4684m</code>
<code>mapreduce.map.memory.mb</code>	2928	2928
<code>mapreduce.reduce.memory.mb</code>	5856	5856
<code>yarn.app.mapreduce.am.resource.mb</code>	5856	5856
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	11712	5856
<code>yarn.nodemanager.resource.memory-mb</code>	11712	5856

## m7a.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## m7a.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.java.opts	-Xmx5428m	-Xmx5428m
mapreduce.map.memory.mb	3392	3392
mapreduce.reduce.memory.mb	6784	6784
yarn.app.mapreduce.am.resource.mb	6784	6784

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

## m7a.8xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx2918m</code>	<code>-Xmx2918m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5836m</code>	<code>-Xmx5836m</code>
<code>mapreduce.map.memory.mb</code>	3648	3648
<code>mapreduce.reduce.memory.mb</code>	7296	7296
<code>yarn.app.mapreduce.am.resource.mb</code>	7296	7296
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184



## m7a.12xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx2986m</code>	<code>-Xmx2986m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5972m</code>	<code>-Xmx5972m</code>
<code>mapreduce.map.memory.mb</code>	3733	3733
<code>mapreduce.reduce.memory.mb</code>	7466	7466
<code>yarn.app.mapreduce.am.resource.mb</code>	7466	7466
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	179200	29880
<code>yarn.nodemanager.resource.memory-mb</code>	179200	29880

## m7a.16xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx3021m</code>	<code>-Xmx3021m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.map.memory.mb</code>	3776	3776
<code>mapreduce.reduce.memory.mb</code>	7552	7552
<code>yarn.app.mapreduce.am.resource.mb</code>	7552	7552

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

## m7a.24xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx3055m</code>	<code>-Xmx3055m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6110m</code>	<code>-Xmx6110m</code>
<code>mapreduce.map.memory.mb</code>	3819	3819
<code>mapreduce.reduce.memory.mb</code>	7638	7638
<code>yarn.app.mapreduce.am.resource.mb</code>	7638	7638
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	366592	30520
<code>yarn.nodemanager.resource.memory-mb</code>	366592	30520

## m7a.32xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx3072m	-Xmx3072m
mapreduce.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.map.memory.mb	3840	3840
mapreduce.reduce.memory.mb	7680	7680
yarn.app.mapreduce.am.resource.mb	7680	7680
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

## m7a.48xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx3089m	-Xmx3089m
mapreduce.java.opts	-Xmx6178m	-Xmx6178m
mapreduce.map.memory.mb	3861	3861
mapreduce.reduce.memory.mb	7722	7722
yarn.app.mapreduce.am.resource.mb	7722	7722

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	741376	30952
<code>yarn.nodemanager.resource.memory-mb</code>	741376	30952

## M7g-Instances

### m7g.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4684m</code>	<code>-Xmx4684m</code>
<code>mapreduce.map.memory.mb</code>	2928	2928
<code>mapreduce.reduce.memory.mb</code>	5856	5856
<code>yarn.app.mapreduce.am.resource.mb</code>	5856	5856
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	11712	5856
<code>yarn.nodemanager.resource.memory-mb</code>	11712	5856

## m7g.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4684m</code>	<code>-Xmx4684m</code>
<code>mapreduce.map.memory.mb</code>	2928	2928
<code>mapreduce.reduce.memory.mb</code>	5856	5856
<code>yarn.app.mapreduce.am.resource.mb</code>	5856	5856
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

## m7g.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx2714m</code>	<code>-Xmx2714m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5428m</code>	<code>-Xmx5428m</code>
<code>mapreduce.map.memory.mb</code>	3392	3392
<code>mapreduce.reduce.memory.mb</code>	6784	6784
<code>yarn.app.mapreduce.am.resource.mb</code>	6784	6784

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

## m7g.8xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx2918m</code>	<code>-Xmx2918m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5836m</code>	<code>-Xmx5836m</code>
<code>mapreduce.map.memory.mb</code>	3648	3648
<code>mapreduce.reduce.memory.mb</code>	7296	7296
<code>yarn.app.mapreduce.am.resource.mb</code>	7296	7296
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

## m7g.12xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx2986m</code>	<code>-Xmx2986m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5972m</code>	<code>-Xmx5972m</code>
<code>mapreduce.map.memory.mb</code>	3733	3733
<code>mapreduce.reduce.memory.mb</code>	7466	7466
<code>yarn.app.mapreduce.am.resource.mb</code>	7466	7466
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	179200	29880
<code>yarn.nodemanager.resource.memory-mb</code>	179200	29880

## m7g.16xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx3021m</code>	<code>-Xmx3021m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.map.memory.mb</code>	3776	3776
<code>mapreduce.reduce.memory.mb</code>	7552	7552
<code>yarn.app.mapreduce.am.resource.mb</code>	7552	7552

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

## M7gd-Instances

### m7gd.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4684m</code>	<code>-Xmx4684m</code>
<code>mapreduce.map.memory.mb</code>	2928	2928
<code>mapreduce.reduce.memory.mb</code>	5856	5856
<code>yarn.app.mapreduce.am.resource.mb</code>	5856	5856
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	11712	5856
<code>yarn.nodemanager.resource.memory-mb</code>	11712	5856



## m7gd.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## m7gd.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.java.opts	-Xmx5428m	-Xmx5428m
mapreduce.map.memory.mb	3392	3392
mapreduce.reduce.memory.mb	6784	6784
yarn.app.mapreduce.am.resource.mb	6784	6784

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

## m7gd.8xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx2918m</code>	<code>-Xmx2918m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5836m</code>	<code>-Xmx5836m</code>
<code>mapreduce.map.memory.mb</code>	3648	3648
<code>mapreduce.reduce.memory.mb</code>	7296	7296
<code>yarn.app.mapreduce.am.resource.mb</code>	7296	7296
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

## m7gd.12xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx2986m	-Xmx2986m
mapreduce.java.opts	-Xmx5972m	-Xmx5972m
mapreduce.map.memory.mb	3733	3733
mapreduce.reduce.memory.mb	7466	7466
yarn.app.mapreduce.am.resource.mb	7466	7466
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	179200	29880
yarn.nodemanager.resource.memory-mb	179200	29880

## m7gd.16xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx3021m	-Xmx3021m
mapreduce.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.map.memory.mb	3776	3776
mapreduce.reduce.memory.mb	7552	7552
yarn.app.mapreduce.am.resource.mb	7552	7552

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

## M7i-Instances

### m7i.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4684m</code>	<code>-Xmx4684m</code>
<code>mapreduce.map.memory.mb</code>	2928	2928
<code>mapreduce.reduce.memory.mb</code>	5856	5856
<code>yarn.app.mapreduce.am.resource.mb</code>	5856	5856
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	11712	5856
<code>yarn.nodemanager.resource.memory-mb</code>	11712	5856

## m7i.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4684m</code>	<code>-Xmx4684m</code>
<code>mapreduce.map.memory.mb</code>	2928	2928
<code>mapreduce.reduce.memory.mb</code>	5856	5856
<code>yarn.app.mapreduce.am.resource.mb</code>	5856	5856
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

## m7i.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx2714m</code>	<code>-Xmx2714m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5428m</code>	<code>-Xmx5428m</code>
<code>mapreduce.map.memory.mb</code>	3392	3392
<code>mapreduce.reduce.memory.mb</code>	6784	6784
<code>yarn.app.mapreduce.am.resource.mb</code>	6784	6784

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

## m7i.8xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx2918m</code>	<code>-Xmx2918m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5836m</code>	<code>-Xmx5836m</code>
<code>mapreduce.map.memory.mb</code>	3648	3648
<code>mapreduce.reduce.memory.mb</code>	7296	7296
<code>yarn.app.mapreduce.am.resource.mb</code>	7296	7296
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

## m7i.12xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx3021m	-Xmx3021m
mapreduce.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.map.memory.mb	3776	3776
mapreduce.reduce.memory.mb	7552	7552
yarn.app.mapreduce.am.resource.mb	7552	7552
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	181248	30208
yarn.nodemanager.resource.memory-mb	181248	30208

## m7i.16xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx3021m	-Xmx3021m
mapreduce.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.map.memory.mb	3776	3776
mapreduce.reduce.memory.mb	7552	7552
yarn.app.mapreduce.am.resource.mb	7552	7552

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

## m7i.24xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx3055m	-Xmx3055m
mapreduce.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.map.memory.mb	3819	3819
mapreduce.reduce.memory.mb	7638	7638
yarn.app.mapreduce.am.resource.mb	7638	7638
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30520
yarn.nodemanager.resource.memory-mb	366592	30520



## m7i.48xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx3089m</code>	<code>-Xmx3089m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6178m</code>	<code>-Xmx6178m</code>
<code>mapreduce.map.memory.mb</code>	3861	3861
<code>mapreduce.reduce.memory.mb</code>	7722	7722
<code>yarn.app.mapreduce.am.resource.mb</code>	7722	7722
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	741376	30952
<code>yarn.nodemanager.resource.memory-mb</code>	741376	30952

## M7i-flex-Instances

## m7i-flex.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4684m</code>	<code>-Xmx4684m</code>
<code>mapreduce.map.memory.mb</code>	2928	2928
<code>mapreduce.reduce.memory.mb</code>	5856	5856

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

## m7i-flex.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## m7i-flex.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx2714m</code>	<code>-Xmx2714m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5428m</code>	<code>-Xmx5428m</code>
<code>mapreduce.map.memory.mb</code>	3392	3392
<code>mapreduce.reduce.memory.mb</code>	6784	6784
<code>yarn.app.mapreduce.am.resource.mb</code>	6784	6784
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

## m7i-flex.8xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx2918m</code>	<code>-Xmx2918m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5836m</code>	<code>-Xmx5836m</code>
<code>mapreduce.map.memory.mb</code>	3648	3648
<code>mapreduce.reduce.memory.mb</code>	7296	7296
<code>yarn.app.mapreduce.am.resource.mb</code>	7296	7296

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

## P2-Instances

### p2.xgroß

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx10854m</code>	<code>-Xmx10854m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx21708m</code>	<code>-Xmx21708m</code>
<code>mapreduce.map.memory.mb</code>	13568	13568
<code>mapreduce.reduce.memory.mb</code>	27136	27136
<code>yarn.app.mapreduce.am.resource.mb</code>	27136	27136
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

## p2.8xgroß

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx12288m</code>	<code>-Xmx12288m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx24576m</code>	<code>-Xmx24576m</code>
<code>mapreduce.map.memory.mb</code>	15360	15360
<code>mapreduce.reduce.memory.mb</code>	30720	30720
<code>yarn.app.mapreduce.am.resource.mb</code>	30720	30720
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	491520	30720
<code>yarn.nodemanager.resource.memory-mb</code>	491520	30720

## p2.16xgroß

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx9267m</code>	<code>-Xmx9267m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx18534m</code>	<code>-Xmx18534m</code>
<code>mapreduce.map.memory.mb</code>	11584	11584
<code>mapreduce.reduce.memory.mb</code>	23168	23168
<code>yarn.app.mapreduce.am.resource.mb</code>	23168	23168

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	741376	23168
<code>yarn.nodemanager.resource.memory-mb</code>	741376	23168

### P3-Instances

#### p3.2xgroß

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx5427m</code>	<code>-Xmx5427m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx10854m</code>	<code>-Xmx10854m</code>
<code>mapreduce.map.memory.mb</code>	6784	6784
<code>mapreduce.reduce.memory.mb</code>	13568	13568
<code>yarn.app.mapreduce.am.resource.mb</code>	13568	13568
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

## p3.8xgroß

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12084m</code>	<code>-Xmx12084m</code>
<code>mapreduce.map.memory.mb</code>	7552	7552
<code>mapreduce.reduce.memory.mb</code>	15104	15104
<code>yarn.app.mapreduce.am.resource.mb</code>	15104	15104
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

## p3.16xgroß

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6144m</code>	<code>-Xmx6144m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12288m</code>	<code>-Xmx12288m</code>
<code>mapreduce.map.memory.mb</code>	7680	7680
<code>mapreduce.reduce.memory.mb</code>	15360	15360
<code>yarn.app.mapreduce.am.resource.mb</code>	15360	15360

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	491520	30720
<code>yarn.nodemanager.resource.memory-mb</code>	491520	30720

## P5-Instances

### p5.48xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx8294m</code>	<code>-Xmx8294m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx16588m</code>	<code>-Xmx16588m</code>
<code>mapreduce.map.memory.mb</code>	10368	10368
<code>mapreduce.reduce.memory.mb</code>	20736	20736
<code>yarn.app.mapreduce.am.resource.mb</code>	20736	20736
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	1990656	20736
<code>yarn.nodemanager.resource.memory-mb</code>	1990656	20736



## R3-Instances

## r3.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## r3.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.java.opts	-Xmx5428m	-Xmx5428m
mapreduce.map.memory.mb	3392	3392
mapreduce.reduce.memory.mb	6784	6784

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.app.mapreduce.am.resource.mb</code>	6784	6784
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

## r3.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx2918m</code>	<code>-Xmx2918m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5836m</code>	<code>-Xmx5836m</code>
<code>mapreduce.map.memory.mb</code>	3648	3648
<code>mapreduce.reduce.memory.mb</code>	7296	7296
<code>yarn.app.mapreduce.am.resource.mb</code>	7296	7296
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

## r3.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx3021m</code>	<code>-Xmx3021m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.map.memory.mb</code>	3776	3776
<code>mapreduce.reduce.memory.mb</code>	7552	7552
<code>yarn.app.mapreduce.am.resource.mb</code>	7552	7552
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

## R4-Instances

## r4.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx4685m</code>	<code>-Xmx4685m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx9370m</code>	<code>-Xmx9370m</code>
<code>mapreduce.map.memory.mb</code>	5856	5856
<code>mapreduce.reduce.memory.mb</code>	11712	11712

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.app.mapreduce.am.resource.mb	11712	11712
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## r4.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx5427m	-Xmx5427m
mapreduce.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.map.memory.mb	6784	6784
mapreduce.reduce.memory.mb	13568	13568
yarn.app.mapreduce.am.resource.mb	13568	13568
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## r4.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx5837m</code>	<code>-Xmx5837m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx11674m</code>	<code>-Xmx11674m</code>
<code>mapreduce.map.memory.mb</code>	7296	7296
<code>mapreduce.reduce.memory.mb</code>	14592	14592
<code>yarn.app.mapreduce.am.resource.mb</code>	14592	14592
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

## r4.8xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12084m</code>	<code>-Xmx12084m</code>
<code>mapreduce.map.memory.mb</code>	7552	7552
<code>mapreduce.reduce.memory.mb</code>	15104	15104
<code>yarn.app.mapreduce.am.resource.mb</code>	15104	15104

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

## r4.16xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6144m</code>	<code>-Xmx6144m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12288m</code>	<code>-Xmx12288m</code>
<code>mapreduce.map.memory.mb</code>	7680	7680
<code>mapreduce.reduce.memory.mb</code>	15360	15360
<code>yarn.app.mapreduce.am.resource.mb</code>	15360	15360
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	491520	30720
<code>yarn.nodemanager.resource.memory-mb</code>	491520	30720

## R5-Instances

## r5.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx4915m	-Xmx4915m
mapreduce.java.opts	-Xmx9830m	-Xmx9830m
mapreduce.map.memory.mb	6144	6144
mapreduce.reduce.memory.mb	12288	12288
yarn.app.mapreduce.am.resource.mb	12288	12288
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	24576	12288
yarn.nodemanager.resource.memory-mb	24576	12288

## r5.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx5734m	-Xmx5734m
mapreduce.java.opts	-Xmx11468m	-Xmx11468m
mapreduce.map.memory.mb	7168	7168
mapreduce.reduce.memory.mb	14336	14336

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.app.mapreduce.am.resource.mb	14336	14336
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	57344	28672
yarn.nodemanager.resource.memory-mb	57344	28672

## r5.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.java.opts	-Xmx12288m	-Xmx12288m
mapreduce.map.memory.mb	7680	7680
mapreduce.reduce.memory.mb	15360	15360
yarn.app.mapreduce.am.resource.mb	15360	15360
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	122880	30720
yarn.nodemanager.resource.memory-mb	122880	30720



## r5.8xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6349m</code>	<code>-Xmx6349m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12698m</code>	<code>-Xmx12698m</code>
<code>mapreduce.map.memory.mb</code>	7936	7936
<code>mapreduce.reduce.memory.mb</code>	15872	15872
<code>yarn.app.mapreduce.am.resource.mb</code>	15872	15872
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	253952	31744
<code>yarn.nodemanager.resource.memory-mb</code>	253952	31744

## r5.12xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6417m</code>	<code>-Xmx6417m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12834m</code>	<code>-Xmx12834m</code>
<code>mapreduce.map.memory.mb</code>	8021	8021
<code>mapreduce.reduce.memory.mb</code>	16042	16042
<code>yarn.app.mapreduce.am.resource.mb</code>	16042	16042

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	385024	32100
<code>yarn.nodemanager.resource.memory-mb</code>	385024	32100

## r5.16xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6451m</code>	<code>-Xmx6451m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12902m</code>	<code>-Xmx12902m</code>
<code>mapreduce.map.memory.mb</code>	8064	8064
<code>mapreduce.reduce.memory.mb</code>	16128	16128
<code>yarn.app.mapreduce.am.resource.mb</code>	16128	16128
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	516096	32256
<code>yarn.nodemanager.resource.memory-mb</code>	516096	32256

## r5.24xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6486m</code>	<code>-Xmx6486m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12972m</code>	<code>-Xmx12972m</code>
<code>mapreduce.map.memory.mb</code>	8107	8107
<code>mapreduce.reduce.memory.mb</code>	16214	16214
<code>yarn.app.mapreduce.am.resource.mb</code>	16214	16214
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	778240	32396
<code>yarn.nodemanager.resource.memory-mb</code>	778240	32396

## R5a-Instances

## r5a.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx4915m</code>	<code>-Xmx4915m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx9830m</code>	<code>-Xmx9830m</code>
<code>mapreduce.map.memory.mb</code>	6144	6144
<code>mapreduce.reduce.memory.mb</code>	12288	12288

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.app.mapreduce.am.resource.mb	12288	12288
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	24576	12288
yarn.nodemanager.resource.memory-mb	24576	12288

## r5a.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx5734m	-Xmx5734m
mapreduce.java.opts	-Xmx11468m	-Xmx11468m
mapreduce.map.memory.mb	7168	7168
mapreduce.reduce.memory.mb	14336	14336
yarn.app.mapreduce.am.resource.mb	14336	14336
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	57344	28672
yarn.nodemanager.resource.memory-mb	57344	28672

## r5a.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6144m</code>	<code>-Xmx6144m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12288m</code>	<code>-Xmx12288m</code>
<code>mapreduce.map.memory.mb</code>	7680	7680
<code>mapreduce.reduce.memory.mb</code>	15360	15360
<code>yarn.app.mapreduce.am.resource.mb</code>	15360	15360
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	122880	30720
<code>yarn.nodemanager.resource.memory-mb</code>	122880	30720

## r5a.8xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6349m</code>	<code>-Xmx6349m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12698m</code>	<code>-Xmx12698m</code>
<code>mapreduce.map.memory.mb</code>	7936	7936
<code>mapreduce.reduce.memory.mb</code>	15872	15872
<code>yarn.app.mapreduce.am.resource.mb</code>	15872	15872

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	253952	31744
yarn.nodemanager.resource.memory-mb	253952	31744

## r5a.12xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx6417m	-Xmx6417m
mapreduce.java.opts	-Xmx12834m	-Xmx12834m
mapreduce.map.memory.mb	8021	8021
mapreduce.reduce.memory.mb	16042	16042
yarn.app.mapreduce.am.resource.mb	16042	16042
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	385024	32100
yarn.nodemanager.resource.memory-mb	385024	32100

## r5a.16xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6451m</code>	<code>-Xmx6451m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12902m</code>	<code>-Xmx12902m</code>
<code>mapreduce.map.memory.mb</code>	8064	8064
<code>mapreduce.reduce.memory.mb</code>	16128	16128
<code>yarn.app.mapreduce.am.resource.mb</code>	16128	16128
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	516096	32256
<code>yarn.nodemanager.resource.memory-mb</code>	516096	32256

## r5a.24xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6486m</code>	<code>-Xmx6486m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12972m</code>	<code>-Xmx12972m</code>
<code>mapreduce.map.memory.mb</code>	8107	8107
<code>mapreduce.reduce.memory.mb</code>	16214	16214
<code>yarn.app.mapreduce.am.resource.mb</code>	16214	16214

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	778240	32396
<code>yarn.nodemanager.resource.memory-mb</code>	778240	32396

## R5ad-Instances

### r5ad.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx4685m</code>	<code>-Xmx4685m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx9370m</code>	<code>-Xmx9370m</code>
<code>mapreduce.map.memory.mb</code>	5856	5856
<code>mapreduce.reduce.memory.mb</code>	11712	11712
<code>yarn.app.mapreduce.am.resource.mb</code>	11712	11712
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712



## r5ad.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx5427m</code>	<code>-Xmx5427m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx10854m</code>	<code>-Xmx10854m</code>
<code>mapreduce.map.memory.mb</code>	6784	6784
<code>mapreduce.reduce.memory.mb</code>	13568	13568
<code>yarn.app.mapreduce.am.resource.mb</code>	13568	13568
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

## r5ad.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx5837m</code>	<code>-Xmx5837m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx11674m</code>	<code>-Xmx11674m</code>
<code>mapreduce.map.memory.mb</code>	7296	7296
<code>mapreduce.reduce.memory.mb</code>	14592	14592
<code>yarn.app.mapreduce.am.resource.mb</code>	14592	14592

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

## r5ad.8xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12084m</code>	<code>-Xmx12084m</code>
<code>mapreduce.map.memory.mb</code>	7552	7552
<code>mapreduce.reduce.memory.mb</code>	15104	15104
<code>yarn.app.mapreduce.am.resource.mb</code>	15104	15104
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

## r5ad.12xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6110m</code>	<code>-Xmx6110m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12220m</code>	<code>-Xmx12220m</code>
<code>mapreduce.map.memory.mb</code>	7637	7637
<code>mapreduce.reduce.memory.mb</code>	15274	15274
<code>yarn.app.mapreduce.am.resource.mb</code>	15274	15274
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	366592	30564
<code>yarn.nodemanager.resource.memory-mb</code>	366592	30564

## r5ad.16xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6246m</code>	<code>-Xmx6246m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12492m</code>	<code>-Xmx12492m</code>
<code>mapreduce.map.memory.mb</code>	7808	7808
<code>mapreduce.reduce.memory.mb</code>	15616	15616
<code>yarn.app.mapreduce.am.resource.mb</code>	15616	15616

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	499712	31232
<code>yarn.nodemanager.resource.memory-mb</code>	499712	31232

## r5ad.24xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6178m</code>	<code>-Xmx6178m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12356m</code>	<code>-Xmx12356m</code>
<code>mapreduce.map.memory.mb</code>	7723	7723
<code>mapreduce.reduce.memory.mb</code>	15446	15446
<code>yarn.app.mapreduce.am.resource.mb</code>	15446	15446
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	741376	30860
<code>yarn.nodemanager.resource.memory-mb</code>	741376	30860

## r5b-Instances

## r5b.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx4685m	-Xmx4685m
mapreduce.java.opts	-Xmx9370m	-Xmx9370m
mapreduce.map.memory.mb	5856	5856
mapreduce.reduce.memory.mb	11712	11712
yarn.app.mapreduce.am.resource.mb	11712	11712
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## r5b.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx5427m	-Xmx5427m
mapreduce.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.map.memory.mb	6784	6784
mapreduce.reduce.memory.mb	13568	13568

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.app.mapreduce.am.resource.mb	13568	13568
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## r5b.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx5837m	-Xmx5837m
mapreduce.java.opts	-Xmx11674m	-Xmx11674m
mapreduce.map.memory.mb	7296	7296
mapreduce.reduce.memory.mb	14592	14592
yarn.app.mapreduce.am.resource.mb	14592	14592
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## r5b.8xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12084m</code>	<code>-Xmx12084m</code>
<code>mapreduce.map.memory.mb</code>	7552	7552
<code>mapreduce.reduce.memory.mb</code>	15104	15104
<code>yarn.app.mapreduce.am.resource.mb</code>	15104	15104
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

## r5b.12xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6110m</code>	<code>-Xmx6110m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12220m</code>	<code>-Xmx12220m</code>
<code>mapreduce.map.memory.mb</code>	7637	7637
<code>mapreduce.reduce.memory.mb</code>	15274	15274
<code>yarn.app.mapreduce.am.resource.mb</code>	15274	15274

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	366592	30564
<code>yarn.nodemanager.resource.memory-mb</code>	366592	30564

## r5b.16xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6144m</code>	<code>-Xmx6144m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12288m</code>	<code>-Xmx12288m</code>
<code>mapreduce.map.memory.mb</code>	7680	7680
<code>mapreduce.reduce.memory.mb</code>	15360	15360
<code>yarn.app.mapreduce.am.resource.mb</code>	15360	15360
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	491520	30720
<code>yarn.nodemanager.resource.memory-mb</code>	491520	30720



## r5b.24xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx6178m	-Xmx6178m
mapreduce.java.opts	-Xmx12356m	-Xmx12356m
mapreduce.map.memory.mb	7723	7723
mapreduce.reduce.memory.mb	15446	15446
yarn.app.mapreduce.am.resource.mb	15446	15446
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	741376	30860
yarn.nodemanager.resource.memory-mb	741376	30860

## R5d-Instances

## r5d.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx4915m	-Xmx4915m
mapreduce.java.opts	-Xmx9830m	-Xmx9830m
mapreduce.map.memory.mb	6144	6144
mapreduce.reduce.memory.mb	12288	12288

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.app.mapreduce.am.resource.mb	12288	12288
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	24576	12288
yarn.nodemanager.resource.memory-mb	24576	12288

## r5d.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx5734m	-Xmx5734m
mapreduce.java.opts	-Xmx11468m	-Xmx11468m
mapreduce.map.memory.mb	7168	7168
mapreduce.reduce.memory.mb	14336	14336
yarn.app.mapreduce.am.resource.mb	14336	14336
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	57344	28672
yarn.nodemanager.resource.memory-mb	57344	28672

## r5d.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6144m</code>	<code>-Xmx6144m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12288m</code>	<code>-Xmx12288m</code>
<code>mapreduce.map.memory.mb</code>	7680	7680
<code>mapreduce.reduce.memory.mb</code>	15360	15360
<code>yarn.app.mapreduce.am.resource.mb</code>	15360	15360
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	122880	30720
<code>yarn.nodemanager.resource.memory-mb</code>	122880	30720

## r5d.8xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6349m</code>	<code>-Xmx6349m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12698m</code>	<code>-Xmx12698m</code>
<code>mapreduce.map.memory.mb</code>	7936	7936
<code>mapreduce.reduce.memory.mb</code>	15872	15872
<code>yarn.app.mapreduce.am.resource.mb</code>	15872	15872

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	253952	31744
<code>yarn.nodemanager.resource.memory-mb</code>	253952	31744

## r5d.12xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6417m</code>	<code>-Xmx6417m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12834m</code>	<code>-Xmx12834m</code>
<code>mapreduce.map.memory.mb</code>	8021	8021
<code>mapreduce.reduce.memory.mb</code>	16042	16042
<code>yarn.app.mapreduce.am.resource.mb</code>	16042	16042
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	385024	32100
<code>yarn.nodemanager.resource.memory-mb</code>	385024	32100

## r5d.16xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx6451m	-Xmx6451m
mapreduce.java.opts	-Xmx12902m	-Xmx12902m
mapreduce.map.memory.mb	8064	8064
mapreduce.reduce.memory.mb	16128	16128
yarn.app.mapreduce.am.resource.mb	16128	16128
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	516096	32256
yarn.nodemanager.resource.memory-mb	516096	32256

## r5d.24xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx6486m	-Xmx6486m
mapreduce.java.opts	-Xmx12972m	-Xmx12972m
mapreduce.map.memory.mb	8107	8107
mapreduce.reduce.memory.mb	16214	16214
yarn.app.mapreduce.am.resource.mb	16214	16214

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	778240	32396
<code>yarn.nodemanager.resource.memory-mb</code>	778240	32396

## R5dn-Instances

### r5dn.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx4685m</code>	<code>-Xmx4685m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx9370m</code>	<code>-Xmx9370m</code>
<code>mapreduce.map.memory.mb</code>	5856	5856
<code>mapreduce.reduce.memory.mb</code>	11712	11712
<code>yarn.app.mapreduce.am.resource.mb</code>	11712	11712
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

## r5dn.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx5427m	-Xmx5427m
mapreduce.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.map.memory.mb	6784	6784
mapreduce.reduce.memory.mb	13568	13568
yarn.app.mapreduce.am.resource.mb	13568	13568
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## r5dn.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx5837m	-Xmx5837m
mapreduce.java.opts	-Xmx11674m	-Xmx11674m
mapreduce.map.memory.mb	7296	7296
mapreduce.reduce.memory.mb	14592	14592
yarn.app.mapreduce.am.resource.mb	14592	14592

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

## r5dn.8xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12084m</code>	<code>-Xmx12084m</code>
<code>mapreduce.map.memory.mb</code>	7552	7552
<code>mapreduce.reduce.memory.mb</code>	15104	15104
<code>yarn.app.mapreduce.am.resource.mb</code>	15104	15104
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208



## r5dn.12xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.java.opts	-Xmx12220m	-Xmx12220m
mapreduce.map.memory.mb	7637	7637
mapreduce.reduce.memory.mb	15274	15274
yarn.app.mapreduce.am.resource.mb	15274	15274
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30564
yarn.nodemanager.resource.memory-mb	366592	30564

## r5dn.16xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.java.opts	-Xmx12288m	-Xmx12288m
mapreduce.map.memory.mb	7680	7680
mapreduce.reduce.memory.mb	15360	15360
yarn.app.mapreduce.am.resource.mb	15360	15360

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	491520	30720
<code>yarn.nodemanager.resource.memory-mb</code>	491520	30720

## r5dn.24xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6178m</code>	<code>-Xmx6178m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12356m</code>	<code>-Xmx12356m</code>
<code>mapreduce.map.memory.mb</code>	7723	7723
<code>mapreduce.reduce.memory.mb</code>	15446	15446
<code>yarn.app.mapreduce.am.resource.mb</code>	15446	15446
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	741376	30860
<code>yarn.nodemanager.resource.memory-mb</code>	741376	30860

## R5n-Instances

## r5n.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx4685m	-Xmx4685m
mapreduce.java.opts	-Xmx9370m	-Xmx9370m
mapreduce.map.memory.mb	5856	5856
mapreduce.reduce.memory.mb	11712	11712
yarn.app.mapreduce.am.resource.mb	11712	11712
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## r5n.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx5427m	-Xmx5427m
mapreduce.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.map.memory.mb	6784	6784
mapreduce.reduce.memory.mb	13568	13568

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.app.mapreduce.am.resource.mb	13568	13568
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## r5n.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx5837m	-Xmx5837m
mapreduce.java.opts	-Xmx11674m	-Xmx11674m
mapreduce.map.memory.mb	7296	7296
mapreduce.reduce.memory.mb	14592	14592
yarn.app.mapreduce.am.resource.mb	14592	14592
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## r5n.8xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12084m</code>	<code>-Xmx12084m</code>
<code>mapreduce.map.memory.mb</code>	7552	7552
<code>mapreduce.reduce.memory.mb</code>	15104	15104
<code>yarn.app.mapreduce.am.resource.mb</code>	15104	15104
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

## r5n.12xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6110m</code>	<code>-Xmx6110m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12220m</code>	<code>-Xmx12220m</code>
<code>mapreduce.map.memory.mb</code>	7637	7637
<code>mapreduce.reduce.memory.mb</code>	15274	15274
<code>yarn.app.mapreduce.am.resource.mb</code>	15274	15274

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	366592	30564
<code>yarn.nodemanager.resource.memory-mb</code>	366592	30564

## r5n.16xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6144m</code>	<code>-Xmx6144m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12288m</code>	<code>-Xmx12288m</code>
<code>mapreduce.map.memory.mb</code>	7680	7680
<code>mapreduce.reduce.memory.mb</code>	15360	15360
<code>yarn.app.mapreduce.am.resource.mb</code>	15360	15360
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	491520	30720
<code>yarn.nodemanager.resource.memory-mb</code>	491520	30720

## r5n.24xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx6178m	-Xmx6178m
mapreduce.java.opts	-Xmx12356m	-Xmx12356m
mapreduce.map.memory.mb	7723	7723
mapreduce.reduce.memory.mb	15446	15446
yarn.app.mapreduce.am.resource.mb	15446	15446
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	741376	30860
yarn.nodemanager.resource.memory-mb	741376	30860

## R6a-Instances

## r6a.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx4685m	-Xmx4685m
mapreduce.java.opts	-Xmx9370m	-Xmx9370m
mapreduce.map.memory.mb	5856	5856
mapreduce.reduce.memory.mb	11712	11712

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.app.mapreduce.am.resource.mb	11712	11712
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## r6a.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx5427m	-Xmx5427m
mapreduce.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.map.memory.mb	6784	6784
mapreduce.reduce.memory.mb	13568	13568
yarn.app.mapreduce.am.resource.mb	13568	13568
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136



## r6a.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx5837m	-Xmx5837m
mapreduce.java.opts	-Xmx11674m	-Xmx11674m
mapreduce.map.memory.mb	7296	7296
mapreduce.reduce.memory.mb	14592	14592
yarn.app.mapreduce.am.resource.mb	14592	14592
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## r6a.8xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.java.opts	-Xmx12084m	-Xmx12084m
mapreduce.map.memory.mb	7552	7552
mapreduce.reduce.memory.mb	15104	15104
yarn.app.mapreduce.am.resource.mb	15104	15104

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

## r6a.12xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6110m</code>	<code>-Xmx6110m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12220m</code>	<code>-Xmx12220m</code>
<code>mapreduce.map.memory.mb</code>	7637	7637
<code>mapreduce.reduce.memory.mb</code>	15274	15274
<code>yarn.app.mapreduce.am.resource.mb</code>	15274	15274
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	366592	30564
<code>yarn.nodemanager.resource.memory-mb</code>	366592	30564

## r6a.16xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.java.opts	-Xmx12288m	-Xmx12288m
mapreduce.map.memory.mb	7680	7680
mapreduce.reduce.memory.mb	15360	15360
yarn.app.mapreduce.am.resource.mb	15360	15360
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

## r6a.24xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx6178m	-Xmx6178m
mapreduce.java.opts	-Xmx12356m	-Xmx12356m
mapreduce.map.memory.mb	7723	7723
mapreduce.reduce.memory.mb	15446	15446
yarn.app.mapreduce.am.resource.mb	15446	15446

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	741376	30860
<code>yarn.nodemanager.resource.memory-mb</code>	741376	30860

## r6a.32xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6195m</code>	<code>-Xmx6195m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12390m</code>	<code>-Xmx12390m</code>
<code>mapreduce.map.memory.mb</code>	7744	7744
<code>mapreduce.reduce.memory.mb</code>	15488	15488
<code>yarn.app.mapreduce.am.resource.mb</code>	15488	15488
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	991232	30976
<code>yarn.nodemanager.resource.memory-mb</code>	991232	30976

## r6a.48xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx6212m	-Xmx6212m
mapreduce.java.opts	-Xmx12424m	-Xmx12424m
mapreduce.map.memory.mb	7765	7765
mapreduce.reduce.memory.mb	15530	15530
yarn.app.mapreduce.am.resource.mb	15530	15530
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	1490944	31124
yarn.nodemanager.resource.memory-mb	1490944	31124

## R6g-Instances

## r6g.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx4685m	-Xmx4685m
mapreduce.java.opts	-Xmx9370m	-Xmx9370m
mapreduce.map.memory.mb	5856	5856
mapreduce.reduce.memory.mb	11712	11712

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.app.mapreduce.am.resource.mb	11712	11712
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## r6g.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx5427m	-Xmx5427m
mapreduce.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.map.memory.mb	6784	6784
mapreduce.reduce.memory.mb	13568	13568
yarn.app.mapreduce.am.resource.mb	13568	13568
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## r6g.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx5837m</code>	<code>-Xmx5837m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx11674m</code>	<code>-Xmx11674m</code>
<code>mapreduce.map.memory.mb</code>	7296	7296
<code>mapreduce.reduce.memory.mb</code>	14592	14592
<code>yarn.app.mapreduce.am.resource.mb</code>	14592	14592
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

## r6g.8xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12084m</code>	<code>-Xmx12084m</code>
<code>mapreduce.map.memory.mb</code>	7552	7552
<code>mapreduce.reduce.memory.mb</code>	15104	15104
<code>yarn.app.mapreduce.am.resource.mb</code>	15104	15104

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

## r6g.12xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6110m</code>	<code>-Xmx6110m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12220m</code>	<code>-Xmx12220m</code>
<code>mapreduce.map.memory.mb</code>	7637	7637
<code>mapreduce.reduce.memory.mb</code>	15274	15274
<code>yarn.app.mapreduce.am.resource.mb</code>	15274	15274
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	366592	30564
<code>yarn.nodemanager.resource.memory-mb</code>	366592	30564



## r6g.16xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.java.opts	-Xmx12288m	-Xmx12288m
mapreduce.map.memory.mb	7680	7680
mapreduce.reduce.memory.mb	15360	15360
yarn.app.mapreduce.am.resource.mb	15360	15360
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

## R6gd-Instances

## r6gd.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx4685m	-Xmx4685m
mapreduce.java.opts	-Xmx9370m	-Xmx9370m
mapreduce.map.memory.mb	5856	5856
mapreduce.reduce.memory.mb	11712	11712

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.app.mapreduce.am.resource.mb	11712	11712
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## r6gd.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx5427m	-Xmx5427m
mapreduce.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.map.memory.mb	6784	6784
mapreduce.reduce.memory.mb	13568	13568
yarn.app.mapreduce.am.resource.mb	13568	13568
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## r6gd.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx5837m	-Xmx5837m
mapreduce.java.opts	-Xmx11674m	-Xmx11674m
mapreduce.map.memory.mb	7296	7296
mapreduce.reduce.memory.mb	14592	14592
yarn.app.mapreduce.am.resource.mb	14592	14592
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## r6gd.8xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.java.opts	-Xmx12084m	-Xmx12084m
mapreduce.map.memory.mb	7552	7552
mapreduce.reduce.memory.mb	15104	15104
yarn.app.mapreduce.am.resource.mb	15104	15104

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

## r6gd.12xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6110m</code>	<code>-Xmx6110m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12220m</code>	<code>-Xmx12220m</code>
<code>mapreduce.map.memory.mb</code>	7637	7637
<code>mapreduce.reduce.memory.mb</code>	15274	15274
<code>yarn.app.mapreduce.am.resource.mb</code>	15274	15274
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	366592	30564
<code>yarn.nodemanager.resource.memory-mb</code>	366592	30564

## r6gd.16xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.java.opts	-Xmx12288m	-Xmx12288m
mapreduce.map.memory.mb	7680	7680
mapreduce.reduce.memory.mb	15360	15360
yarn.app.mapreduce.am.resource.mb	15360	15360
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

## R6i-Instances

## r6i.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx4685m	-Xmx4685m
mapreduce.java.opts	-Xmx9370m	-Xmx9370m
mapreduce.map.memory.mb	5856	5856
mapreduce.reduce.memory.mb	11712	11712

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.app.mapreduce.am.resource.mb	11712	11712
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## r6i.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx5427m	-Xmx5427m
mapreduce.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.map.memory.mb	6784	6784
mapreduce.reduce.memory.mb	13568	13568
yarn.app.mapreduce.am.resource.mb	13568	13568
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## r6i.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx5837m	-Xmx5837m
mapreduce.java.opts	-Xmx11674m	-Xmx11674m
mapreduce.map.memory.mb	7296	7296
mapreduce.reduce.memory.mb	14592	14592
yarn.app.mapreduce.am.resource.mb	14592	14592
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## r6i.8xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.java.opts	-Xmx12084m	-Xmx12084m
mapreduce.map.memory.mb	7552	7552
mapreduce.reduce.memory.mb	15104	15104
yarn.app.mapreduce.am.resource.mb	15104	15104

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

## r6i.12xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6110m</code>	<code>-Xmx6110m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12220m</code>	<code>-Xmx12220m</code>
<code>mapreduce.map.memory.mb</code>	7637	7637
<code>mapreduce.reduce.memory.mb</code>	15274	15274
<code>yarn.app.mapreduce.am.resource.mb</code>	15274	15274
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	366592	30564
<code>yarn.nodemanager.resource.memory-mb</code>	366592	30564



## r6i.16xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6144m</code>	<code>-Xmx6144m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12288m</code>	<code>-Xmx12288m</code>
<code>mapreduce.map.memory.mb</code>	7680	7680
<code>mapreduce.reduce.memory.mb</code>	15360	15360
<code>yarn.app.mapreduce.am.resource.mb</code>	15360	15360
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	491520	30720
<code>yarn.nodemanager.resource.memory-mb</code>	491520	30720

## r6i.24xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6178m</code>	<code>-Xmx6178m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12356m</code>	<code>-Xmx12356m</code>
<code>mapreduce.map.memory.mb</code>	7723	7723
<code>mapreduce.reduce.memory.mb</code>	15446	15446
<code>yarn.app.mapreduce.am.resource.mb</code>	15446	15446

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	741376	30860
<code>yarn.nodemanager.resource.memory-mb</code>	741376	30860

## r6i.32xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6029m</code>	<code>-Xmx6029m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12058m</code>	<code>-Xmx12058m</code>
<code>mapreduce.map.memory.mb</code>	7536	7536
<code>mapreduce.reduce.memory.mb</code>	15072	15072
<code>yarn.app.mapreduce.am.resource.mb</code>	15072	15072
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	964608	30144
<code>yarn.nodemanager.resource.memory-mb</code>	964608	30144

## R6id-Instances

## r6id.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx4685m	-Xmx4685m
mapreduce.java.opts	-Xmx9370m	-Xmx9370m
mapreduce.map.memory.mb	5856	5856
mapreduce.reduce.memory.mb	11712	11712
yarn.app.mapreduce.am.resource.mb	11712	11712
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## r6id.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx5427m	-Xmx5427m
mapreduce.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.map.memory.mb	6784	6784
mapreduce.reduce.memory.mb	13568	13568

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.app.mapreduce.am.resource.mb	13568	13568
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## r6id.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx5837m	-Xmx5837m
mapreduce.java.opts	-Xmx11674m	-Xmx11674m
mapreduce.map.memory.mb	7296	7296
mapreduce.reduce.memory.mb	14592	14592
yarn.app.mapreduce.am.resource.mb	14592	14592
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## r6id.8xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12084m</code>	<code>-Xmx12084m</code>
<code>mapreduce.map.memory.mb</code>	7552	7552
<code>mapreduce.reduce.memory.mb</code>	15104	15104
<code>yarn.app.mapreduce.am.resource.mb</code>	15104	15104
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

## r6id.12xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6110m</code>	<code>-Xmx6110m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12220m</code>	<code>-Xmx12220m</code>
<code>mapreduce.map.memory.mb</code>	7637	7637
<code>mapreduce.reduce.memory.mb</code>	15274	15274
<code>yarn.app.mapreduce.am.resource.mb</code>	15274	15274

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30564
yarn.nodemanager.resource.memory-mb	366592	30564

## r6id.16xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.java.opts	-Xmx12288m	-Xmx12288m
mapreduce.map.memory.mb	7680	7680
mapreduce.reduce.memory.mb	15360	15360
yarn.app.mapreduce.am.resource.mb	15360	15360
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

## r6id.24xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6178m</code>	<code>-Xmx6178m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12356m</code>	<code>-Xmx12356m</code>
<code>mapreduce.map.memory.mb</code>	7723	7723
<code>mapreduce.reduce.memory.mb</code>	15446	15446
<code>yarn.app.mapreduce.am.resource.mb</code>	15446	15446
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	741376	30860
<code>yarn.nodemanager.resource.memory-mb</code>	741376	30860

## r6id.32xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6195m</code>	<code>-Xmx6195m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12390m</code>	<code>-Xmx12390m</code>
<code>mapreduce.map.memory.mb</code>	7744	7744
<code>mapreduce.reduce.memory.mb</code>	15488	15488
<code>yarn.app.mapreduce.am.resource.mb</code>	15488	15488

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	991232	30976
<code>yarn.nodemanager.resource.memory-mb</code>	991232	30976

## R6idn-Instances

### r6idn.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx4685m</code>	<code>-Xmx4685m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx9370m</code>	<code>-Xmx9370m</code>
<code>mapreduce.map.memory.mb</code>	5856	5856
<code>mapreduce.reduce.memory.mb</code>	11712	11712
<code>yarn.app.mapreduce.am.resource.mb</code>	11712	11712
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712



## r6idn.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx5427m</code>	<code>-Xmx5427m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx10854m</code>	<code>-Xmx10854m</code>
<code>mapreduce.map.memory.mb</code>	6784	6784
<code>mapreduce.reduce.memory.mb</code>	13568	13568
<code>yarn.app.mapreduce.am.resource.mb</code>	13568	13568
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

## r6idn.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx5837m</code>	<code>-Xmx5837m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx11674m</code>	<code>-Xmx11674m</code>
<code>mapreduce.map.memory.mb</code>	7296	7296
<code>mapreduce.reduce.memory.mb</code>	14592	14592
<code>yarn.app.mapreduce.am.resource.mb</code>	14592	14592

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

## r6idn.8xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12084m</code>	<code>-Xmx12084m</code>
<code>mapreduce.map.memory.mb</code>	7552	7552
<code>mapreduce.reduce.memory.mb</code>	15104	15104
<code>yarn.app.mapreduce.am.resource.mb</code>	15104	15104
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

## r6idn.12xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6110m</code>	<code>-Xmx6110m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12220m</code>	<code>-Xmx12220m</code>
<code>mapreduce.map.memory.mb</code>	7637	7637
<code>mapreduce.reduce.memory.mb</code>	15274	15274
<code>yarn.app.mapreduce.am.resource.mb</code>	15274	15274
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	366592	30564
<code>yarn.nodemanager.resource.memory-mb</code>	366592	30564

## r6idn.16xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6144m</code>	<code>-Xmx6144m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12288m</code>	<code>-Xmx12288m</code>
<code>mapreduce.map.memory.mb</code>	7680	7680
<code>mapreduce.reduce.memory.mb</code>	15360	15360
<code>yarn.app.mapreduce.am.resource.mb</code>	15360	15360

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

## r6idn.24xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx6178m	-Xmx6178m
mapreduce.java.opts	-Xmx12356m	-Xmx12356m
mapreduce.map.memory.mb	7723	7723
mapreduce.reduce.memory.mb	15446	15446
yarn.app.mapreduce.am.resource.mb	15446	15446
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	741376	30860
yarn.nodemanager.resource.memory-mb	741376	30860

## r6idn.32xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx6195m	-Xmx6195m
mapreduce.java.opts	-Xmx12390m	-Xmx12390m
mapreduce.map.memory.mb	7744	7744
mapreduce.reduce.memory.mb	15488	15488
yarn.app.mapreduce.am.resource.mb	15488	15488
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	991232	30976
yarn.nodemanager.resource.memory-mb	991232	30976

## R6in-Instances

## r6in.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx4685m	-Xmx4685m
mapreduce.java.opts	-Xmx9370m	-Xmx9370m
mapreduce.map.memory.mb	5856	5856
mapreduce.reduce.memory.mb	11712	11712

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.app.mapreduce.am.resource.mb	11712	11712
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## r6in.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx5427m	-Xmx5427m
mapreduce.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.map.memory.mb	6784	6784
mapreduce.reduce.memory.mb	13568	13568
yarn.app.mapreduce.am.resource.mb	13568	13568
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## r6in.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx5837m</code>	<code>-Xmx5837m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx11674m</code>	<code>-Xmx11674m</code>
<code>mapreduce.map.memory.mb</code>	7296	7296
<code>mapreduce.reduce.memory.mb</code>	14592	14592
<code>yarn.app.mapreduce.am.resource.mb</code>	14592	14592
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

## r6in.8xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12084m</code>	<code>-Xmx12084m</code>
<code>mapreduce.map.memory.mb</code>	7552	7552
<code>mapreduce.reduce.memory.mb</code>	15104	15104
<code>yarn.app.mapreduce.am.resource.mb</code>	15104	15104

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

## r6in.12xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6110m</code>	<code>-Xmx6110m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12220m</code>	<code>-Xmx12220m</code>
<code>mapreduce.map.memory.mb</code>	7637	7637
<code>mapreduce.reduce.memory.mb</code>	15274	15274
<code>yarn.app.mapreduce.am.resource.mb</code>	15274	15274
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	366592	30564
<code>yarn.nodemanager.resource.memory-mb</code>	366592	30564



## r6in.16xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.java.opts	-Xmx12288m	-Xmx12288m
mapreduce.map.memory.mb	7680	7680
mapreduce.reduce.memory.mb	15360	15360
yarn.app.mapreduce.am.resource.mb	15360	15360
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

## r6in.24xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx6178m	-Xmx6178m
mapreduce.java.opts	-Xmx12356m	-Xmx12356m
mapreduce.map.memory.mb	7723	7723
mapreduce.reduce.memory.mb	15446	15446
yarn.app.mapreduce.am.resource.mb	15446	15446

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	741376	30860
<code>yarn.nodemanager.resource.memory-mb</code>	741376	30860

## r6in.32xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6195m</code>	<code>-Xmx6195m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12390m</code>	<code>-Xmx12390m</code>
<code>mapreduce.map.memory.mb</code>	7744	7744
<code>mapreduce.reduce.memory.mb</code>	15488	15488
<code>yarn.app.mapreduce.am.resource.mb</code>	15488	15488
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	991232	30976
<code>yarn.nodemanager.resource.memory-mb</code>	991232	30976

## R7a-Instances

## r7a.x groß

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx4685m	-Xmx4685m
mapreduce.java.opts	-Xmx9370m	-Xmx9370m
mapreduce.map.memory.mb	5856	5856
mapreduce.reduce.memory.mb	11712	11712
yarn.app.mapreduce.am.resource.mb	11712	11712
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## r7a.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx5427m	-Xmx5427m
mapreduce.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.map.memory.mb	6784	6784
mapreduce.reduce.memory.mb	13568	13568

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.app.mapreduce.am.resource.mb	13568	13568
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## r7a.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx5837m	-Xmx5837m
mapreduce.java.opts	-Xmx11674m	-Xmx11674m
mapreduce.map.memory.mb	7296	7296
mapreduce.reduce.memory.mb	14592	14592
yarn.app.mapreduce.am.resource.mb	14592	14592
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## r7a.8xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.java.opts	-Xmx12084m	-Xmx12084m
mapreduce.map.memory.mb	7552	7552
mapreduce.reduce.memory.mb	15104	15104
yarn.app.mapreduce.am.resource.mb	15104	15104
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

## r7a.12xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.java.opts	-Xmx12220m	-Xmx12220m
mapreduce.map.memory.mb	7637	7637
mapreduce.reduce.memory.mb	15274	15274
yarn.app.mapreduce.am.resource.mb	15274	15274

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	366592	30564
<code>yarn.nodemanager.resource.memory-mb</code>	366592	30564

## r7a.16xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6144m</code>	<code>-Xmx6144m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12288m</code>	<code>-Xmx12288m</code>
<code>mapreduce.map.memory.mb</code>	7680	7680
<code>mapreduce.reduce.memory.mb</code>	15360	15360
<code>yarn.app.mapreduce.am.resource.mb</code>	15360	15360
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	491520	30720
<code>yarn.nodemanager.resource.memory-mb</code>	491520	30720

## r7a.24xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6178m</code>	<code>-Xmx6178m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12356m</code>	<code>-Xmx12356m</code>
<code>mapreduce.map.memory.mb</code>	7723	7723
<code>mapreduce.reduce.memory.mb</code>	15446	15446
<code>yarn.app.mapreduce.am.resource.mb</code>	15446	15446
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	741376	30860
<code>yarn.nodemanager.resource.memory-mb</code>	741376	30860

## r7a.32xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6195m</code>	<code>-Xmx6195m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12390m</code>	<code>-Xmx12390m</code>
<code>mapreduce.map.memory.mb</code>	7744	7744
<code>mapreduce.reduce.memory.mb</code>	15488	15488
<code>yarn.app.mapreduce.am.resource.mb</code>	15488	15488

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	991232	30976
yarn.nodemanager.resource.memory-mb	991232	30976

## r7a.48xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx6212m	-Xmx6212m
mapreduce.java.opts	-Xmx12424m	-Xmx12424m
mapreduce.map.memory.mb	7765	7765
mapreduce.reduce.memory.mb	15530	15530
yarn.app.mapreduce.am.resource.mb	15530	15530
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	1490944	31124
yarn.nodemanager.resource.memory-mb	1490944	31124



## R7g-Instances

## r7g.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx4685m	-Xmx4685m
mapreduce.java.opts	-Xmx9370m	-Xmx9370m
mapreduce.map.memory.mb	5856	5856
mapreduce.reduce.memory.mb	11712	11712
yarn.app.mapreduce.am.resource.mb	11712	11712
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## r7g.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx5427m	-Xmx5427m
mapreduce.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.map.memory.mb	6784	6784
mapreduce.reduce.memory.mb	13568	13568

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.app.mapreduce.am.resource.mb	13568	13568
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## r7g.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx5837m	-Xmx5837m
mapreduce.java.opts	-Xmx11674m	-Xmx11674m
mapreduce.map.memory.mb	7296	7296
mapreduce.reduce.memory.mb	14592	14592
yarn.app.mapreduce.am.resource.mb	14592	14592
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## r7g.8xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12084m</code>	<code>-Xmx12084m</code>
<code>mapreduce.map.memory.mb</code>	7552	7552
<code>mapreduce.reduce.memory.mb</code>	15104	15104
<code>yarn.app.mapreduce.am.resource.mb</code>	15104	15104
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

## r7g.12xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6110m</code>	<code>-Xmx6110m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12220m</code>	<code>-Xmx12220m</code>
<code>mapreduce.map.memory.mb</code>	7637	7637
<code>mapreduce.reduce.memory.mb</code>	15274	15274
<code>yarn.app.mapreduce.am.resource.mb</code>	15274	15274

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30564
yarn.nodemanager.resource.memory-mb	366592	30564

## r7g.16xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.java.opts	-Xmx12288m	-Xmx12288m
mapreduce.map.memory.mb	7680	7680
mapreduce.reduce.memory.mb	15360	15360
yarn.app.mapreduce.am.resource.mb	15360	15360
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

## R7gd-Instances

## r7gd.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx4685m	-Xmx4685m
mapreduce.java.opts	-Xmx9370m	-Xmx9370m
mapreduce.map.memory.mb	5856	5856
mapreduce.reduce.memory.mb	11712	11712
yarn.app.mapreduce.am.resource.mb	11712	11712
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## r7gd.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx5427m	-Xmx5427m
mapreduce.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.map.memory.mb	6784	6784
mapreduce.reduce.memory.mb	13568	13568

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.app.mapreduce.am.resource.mb	13568	13568
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## r7gd.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx5837m	-Xmx5837m
mapreduce.java.opts	-Xmx11674m	-Xmx11674m
mapreduce.map.memory.mb	7296	7296
mapreduce.reduce.memory.mb	14592	14592
yarn.app.mapreduce.am.resource.mb	14592	14592
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## r7gd.8xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12084m</code>	<code>-Xmx12084m</code>
<code>mapreduce.map.memory.mb</code>	7552	7552
<code>mapreduce.reduce.memory.mb</code>	15104	15104
<code>yarn.app.mapreduce.am.resource.mb</code>	15104	15104
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

## r7gd.12xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6110m</code>	<code>-Xmx6110m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12220m</code>	<code>-Xmx12220m</code>
<code>mapreduce.map.memory.mb</code>	7637	7637
<code>mapreduce.reduce.memory.mb</code>	15274	15274
<code>yarn.app.mapreduce.am.resource.mb</code>	15274	15274

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30564
yarn.nodemanager.resource.memory-mb	366592	30564

## r7gd.16xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.java.opts	-Xmx12288m	-Xmx12288m
mapreduce.map.memory.mb	7680	7680
mapreduce.reduce.memory.mb	15360	15360
yarn.app.mapreduce.am.resource.mb	15360	15360
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720



## r7i-Instances

## r7i.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx4685m	-Xmx4685m
mapreduce.java.opts	-Xmx9370m	-Xmx9370m
mapreduce.map.memory.mb	5856	5856
mapreduce.reduce.memory.mb	11712	11712
yarn.app.mapreduce.am.resource.mb	11712	11712
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## r7i.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx5427m	-Xmx5427m
mapreduce.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.map.memory.mb	6784	6784
mapreduce.reduce.memory.mb	13568	13568

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.app.mapreduce.am.resource.mb	13568	13568
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## r7i.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx5837m	-Xmx5837m
mapreduce.java.opts	-Xmx11674m	-Xmx11674m
mapreduce.map.memory.mb	7296	7296
mapreduce.reduce.memory.mb	14592	14592
yarn.app.mapreduce.am.resource.mb	14592	14592
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## r7i.8xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12084m</code>	<code>-Xmx12084m</code>
<code>mapreduce.map.memory.mb</code>	7552	7552
<code>mapreduce.reduce.memory.mb</code>	15104	15104
<code>yarn.app.mapreduce.am.resource.mb</code>	15104	15104
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

## r7i.12xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6110m</code>	<code>-Xmx6110m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12220m</code>	<code>-Xmx12220m</code>
<code>mapreduce.map.memory.mb</code>	7637	7637
<code>mapreduce.reduce.memory.mb</code>	15274	15274
<code>yarn.app.mapreduce.am.resource.mb</code>	15274	15274

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	366592	30564
<code>yarn.nodemanager.resource.memory-mb</code>	366592	30564

## r7i.16xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6144m</code>	<code>-Xmx6144m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12288m</code>	<code>-Xmx12288m</code>
<code>mapreduce.map.memory.mb</code>	7680	7680
<code>mapreduce.reduce.memory.mb</code>	15360	15360
<code>yarn.app.mapreduce.am.resource.mb</code>	15360	15360
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	491520	30720
<code>yarn.nodemanager.resource.memory-mb</code>	491520	30720

## r7i.24xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6178m</code>	<code>-Xmx6178m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12356m</code>	<code>-Xmx12356m</code>
<code>mapreduce.map.memory.mb</code>	7723	7723
<code>mapreduce.reduce.memory.mb</code>	15446	15446
<code>yarn.app.mapreduce.am.resource.mb</code>	15446	15446
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	741376	30860
<code>yarn.nodemanager.resource.memory-mb</code>	741376	30860

## r7i.48xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6212m</code>	<code>-Xmx6212m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12424m</code>	<code>-Xmx12424m</code>
<code>mapreduce.map.memory.mb</code>	7765	7765
<code>mapreduce.reduce.memory.mb</code>	15530	15530
<code>yarn.app.mapreduce.am.resource.mb</code>	15530	15530

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	1490944	31124
<code>yarn.nodemanager.resource.memory-mb</code>	1490944	31124

## R7iz-Instances

### r7iz.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx4685m</code>	<code>-Xmx4685m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx9370m</code>	<code>-Xmx9370m</code>
<code>mapreduce.map.memory.mb</code>	5856	5856
<code>mapreduce.reduce.memory.mb</code>	11712	11712
<code>yarn.app.mapreduce.am.resource.mb</code>	11712	11712
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

## r7iz.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx5427m	-Xmx5427m
mapreduce.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.map.memory.mb	6784	6784
mapreduce.reduce.memory.mb	13568	13568
yarn.app.mapreduce.am.resource.mb	13568	13568
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## r7iz.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx5837m	-Xmx5837m
mapreduce.java.opts	-Xmx11674m	-Xmx11674m
mapreduce.map.memory.mb	7296	7296
mapreduce.reduce.memory.mb	14592	14592
yarn.app.mapreduce.am.resource.mb	14592	14592

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## r7iz.8xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.java.opts	-Xmx12084m	-Xmx12084m
mapreduce.map.memory.mb	7552	7552
mapreduce.reduce.memory.mb	15104	15104
yarn.app.mapreduce.am.resource.mb	15104	15104
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208



## r7iz.12xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.java.opts	-Xmx12220m	-Xmx12220m
mapreduce.map.memory.mb	7637	7637
mapreduce.reduce.memory.mb	15274	15274
yarn.app.mapreduce.am.resource.mb	15274	15274
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30564
yarn.nodemanager.resource.memory-mb	366592	30564

## r7iz.16xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.java.opts	-Xmx12288m	-Xmx12288m
mapreduce.map.memory.mb	7680	7680
mapreduce.reduce.memory.mb	15360	15360
yarn.app.mapreduce.am.resource.mb	15360	15360

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	491520	30720
<code>yarn.nodemanager.resource.memory-mb</code>	491520	30720

## r7iz.32xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6029m</code>	<code>-Xmx6029m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12058m</code>	<code>-Xmx12058m</code>
<code>mapreduce.map.memory.mb</code>	7536	7536
<code>mapreduce.reduce.memory.mb</code>	15072	15072
<code>yarn.app.mapreduce.am.resource.mb</code>	15072	15072
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	964608	30144
<code>yarn.nodemanager.resource.memory-mb</code>	964608	30144

## X1-Instances

## x1.16xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx12058m	-Xmx12058m
mapreduce.java.opts	-Xmx24116m	-Xmx24116m
mapreduce.map.memory.mb	15072	15072
mapreduce.reduce.memory.mb	30144	30144
yarn.app.mapreduce.am.resource.mb	30144	30144
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	964608	30144
yarn.nodemanager.resource.memory-mb	964608	30144

## x1.32xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx12109m	-Xmx12109m
mapreduce.java.opts	-Xmx24218m	-Xmx24218m
mapreduce.map.memory.mb	15136	15136
mapreduce.reduce.memory.mb	30272	30272

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.app.mapreduce.am.resource.mb	30272	30272
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	1937408	30272
yarn.nodemanager.resource.memory-mb	1937408	30272

## X1e-Instances

### x1e.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx22682m	-Xmx22682m
mapreduce.java.opts	-Xmx45364m	-Xmx45364m
mapreduce.map.memory.mb	28352	28352
mapreduce.reduce.memory.mb	56704	56704
yarn.app.mapreduce.am.resource.mb	56704	56704
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	113408	0

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.nodemanager.resource.memory-mb</code>	113408	0

## x1e.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx23501m</code>	<code>-Xmx23501m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx47002m</code>	<code>-Xmx47002m</code>
<code>mapreduce.map.memory.mb</code>	29376	29376
<code>mapreduce.reduce.memory.mb</code>	58752	58752
<code>yarn.app.mapreduce.am.resource.mb</code>	58752	58752
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	235008	0
<code>yarn.nodemanager.resource.memory-mb</code>	235008	0

## x1e.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx23910m</code>	<code>-Xmx23910m</code>

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.java.opts	-Xmx47820m	-Xmx47820m
mapreduce.map.memory.mb	29888	29888
mapreduce.reduce.memory.mb	59776	59776
yarn.app.mapreduce.am.resource.mb	59776	59776
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	478208	0
yarn.nodemanager.resource.memory-mb	478208	0

## x1e.8xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx24115m	-Xmx24115m
mapreduce.java.opts	-Xmx48230m	-Xmx48230m
mapreduce.map.memory.mb	30144	30144
mapreduce.reduce.memory.mb	60288	60288
yarn.app.mapreduce.am.resource.mb	60288	60288
yarn.scheduler.minimum-allocation-mb	32	32

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.maximum-allocation-mb</code>	964608	0
<code>yarn.nodemanager.resource-memory-mb</code>	964608	0

## x1e.16xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx24218m</code>	<code>-Xmx24218m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx48436m</code>	<code>-Xmx48436m</code>
<code>mapreduce.map.memory.mb</code>	30272	30272
<code>mapreduce.reduce.memory.mb</code>	60544	60544
<code>yarn.app.mapreduce.am.resource.mb</code>	60544	60544
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	1937408	0
<code>yarn.nodemanager.resource-memory-mb</code>	1937408	0

## x1e.32xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx24269m	-Xmx24269m
mapreduce.java.opts	-Xmx48538m	-Xmx48538m
mapreduce.map.memory.mb	30336	30336
mapreduce.reduce.memory.mb	60672	60672
yarn.app.mapreduce.am.resource.mb	60672	60672
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	3883008	0
yarn.nodemanager.resource.memory-mb	3883008	0

## X2gd-Instances

## x2gd.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.java.opts	-Xmx21708m	-Xmx21708m
mapreduce.map.memory.mb	13568	13568
mapreduce.reduce.memory.mb	27136	27136



Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.app.mapreduce.am.resource.mb	27136	27136
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## x2gd.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx11674m	-Xmx11674m
mapreduce.java.opts	-Xmx23348m	-Xmx23348m
mapreduce.map.memory.mb	14592	14592
mapreduce.reduce.memory.mb	29184	29184
yarn.app.mapreduce.am.resource.mb	29184	29184
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## x2gd.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx12083m</code>	<code>-Xmx12083m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx24166m</code>	<code>-Xmx24166m</code>
<code>mapreduce.map.memory.mb</code>	15104	15104
<code>mapreduce.reduce.memory.mb</code>	30208	30208
<code>yarn.app.mapreduce.am.resource.mb</code>	30208	30208
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

## x2gd.8xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx12288m</code>	<code>-Xmx12288m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx24576m</code>	<code>-Xmx24576m</code>
<code>mapreduce.map.memory.mb</code>	15360	15360
<code>mapreduce.reduce.memory.mb</code>	30720	30720
<code>yarn.app.mapreduce.am.resource.mb</code>	30720	30720

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

## x2gd.12xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx12356m	-Xmx12356m
mapreduce.java.opts	-Xmx24712m	-Xmx24712m
mapreduce.map.memory.mb	15445	15445
mapreduce.reduce.memory.mb	30890	30890
yarn.app.mapreduce.am.resource.mb	30890	30890
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	741376	30906
yarn.nodemanager.resource.memory-mb	741376	30906

## x2gd.16xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx12390m</code>	<code>-Xmx12390m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx24780m</code>	<code>-Xmx24780m</code>
<code>mapreduce.map.memory.mb</code>	15488	15488
<code>mapreduce.reduce.memory.mb</code>	30976	30976
<code>yarn.app.mapreduce.am.resource.mb</code>	30976	30976
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	991232	30976
<code>yarn.nodemanager.resource.memory-mb</code>	991232	30976

## x2idn-Instances

## x2idn.16xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx12390m</code>	<code>-Xmx12390m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx24780m</code>	<code>-Xmx24780m</code>
<code>mapreduce.map.memory.mb</code>	15488	15488
<code>mapreduce.reduce.memory.mb</code>	30976	30976

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.app.mapreduce.am.resource.mb	30976	30976
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	991232	30976
yarn.nodemanager.resource.memory-mb	991232	30976

## x2idn.24xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx12425m	-Xmx12425m
mapreduce.java.opts	-Xmx24850m	-Xmx24850m
mapreduce.map.memory.mb	15531	15531
mapreduce.reduce.memory.mb	31062	31062
yarn.app.mapreduce.am.resource.mb	31062	31062
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	1490944	31030
yarn.nodemanager.resource.memory-mb	1490944	31030

## x2idn.32xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx12442m	-Xmx12442m
mapreduce.java.opts	-Xmx24884m	-Xmx24884m
mapreduce.map.memory.mb	15552	15552
mapreduce.reduce.memory.mb	31104	31104
yarn.app.mapreduce.am.resource.mb	31104	31104
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	1990656	31104
yarn.nodemanager.resource.memory-mb	1990656	31104

## X2iedn-Instances

## x2iedn.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx23347m	-Xmx23347m
mapreduce.java.opts	-Xmx46694m	-Xmx46694m
mapreduce.map.memory.mb	29184	29184
mapreduce.reduce.memory.mb	58368	58368

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.app.mapreduce.am.resource.mb	58368	58368
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	0
yarn.nodemanager.resource.memory-mb	116736	0

## x2iedn.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx24166m	-Xmx24166m
mapreduce.java.opts	-Xmx48332m	-Xmx48332m
mapreduce.map.memory.mb	30208	30208
mapreduce.reduce.memory.mb	60416	60416
yarn.app.mapreduce.am.resource.mb	60416	60416
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	0
yarn.nodemanager.resource.memory-mb	241664	0

## x2iedn.4xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx24576m</code>	<code>-Xmx24576m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx49152m</code>	<code>-Xmx49152m</code>
<code>mapreduce.map.memory.mb</code>	30720	30720
<code>mapreduce.reduce.memory.mb</code>	61440	61440
<code>yarn.app.mapreduce.am.resource.mb</code>	61440	61440
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	491520	0
<code>yarn.nodemanager.resource.memory-mb</code>	491520	0

## x2iedn.8xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx24781m</code>	<code>-Xmx24781m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx49562m</code>	<code>-Xmx49562m</code>
<code>mapreduce.map.memory.mb</code>	30976	30976
<code>mapreduce.reduce.memory.mb</code>	61952	61952
<code>yarn.app.mapreduce.am.resource.mb</code>	61952	61952



Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	991232	0
<code>yarn.nodemanager.resource.memory-mb</code>	991232	0

## x2iedn.16xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx24883m</code>	<code>-Xmx24883m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx49766m</code>	<code>-Xmx49766m</code>
<code>mapreduce.map.memory.mb</code>	31104	31104
<code>mapreduce.reduce.memory.mb</code>	62208	62208
<code>yarn.app.mapreduce.am.resource.mb</code>	62208	62208
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	1990656	0
<code>yarn.nodemanager.resource.memory-mb</code>	1990656	0

## x2iedn.24xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx24918m</code>	<code>-Xmx24918m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx49836m</code>	<code>-Xmx49836m</code>
<code>mapreduce.map.memory.mb</code>	31147	31147
<code>mapreduce.reduce.memory.mb</code>	62294	62294
<code>yarn.app.mapreduce.am.resource.mb</code>	62294	62294
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	2990080	-32
<code>yarn.nodemanager.resource.memory-mb</code>	2990080	-32

## x2iedn.32xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx24934m</code>	<code>-Xmx24934m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx49868m</code>	<code>-Xmx49868m</code>
<code>mapreduce.map.memory.mb</code>	31168	31168
<code>mapreduce.reduce.memory.mb</code>	62336	62336
<code>yarn.app.mapreduce.am.resource.mb</code>	62336	62336

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	3989504	0
<code>yarn.nodemanager.resource.memory-mb</code>	3989504	0

## z1d-Instances

## z1d.xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx4915m</code>	<code>-Xmx4915m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx9830m</code>	<code>-Xmx9830m</code>
<code>mapreduce.map.memory.mb</code>	6144	6144
<code>mapreduce.reduce.memory.mb</code>	12288	12288
<code>yarn.app.mapreduce.am.resource.mb</code>	12288	12288
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	24576	12288
<code>yarn.nodemanager.resource.memory-mb</code>	24576	12288

## z1d.2xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx5734m</code>	<code>-Xmx5734m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx11468m</code>	<code>-Xmx11468m</code>
<code>mapreduce.map.memory.mb</code>	7168	7168
<code>mapreduce.reduce.memory.mb</code>	14336	14336
<code>yarn.app.mapreduce.am.resource.mb</code>	14336	14336
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	57344	28672
<code>yarn.nodemanager.resource.memory-mb</code>	57344	28672

## z1d.3xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
<code>mapreduce.map.java.opts</code>	<code>-Xmx6007m</code>	<code>-Xmx6007m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12014m</code>	<code>-Xmx12014m</code>
<code>mapreduce.map.memory.mb</code>	7509	7509
<code>mapreduce.reduce.memory.mb</code>	15018	15018
<code>yarn.app.mapreduce.am.resource.mb</code>	15018	15018

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	90112	30040
yarn.nodemanager.resource.memory-mb	90112	30040

## z1d.6xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx6281m	-Xmx6281m
mapreduce.java.opts	-Xmx12562m	-Xmx12562m
mapreduce.map.memory.mb	7851	7851
mapreduce.reduce.memory.mb	15702	15702
yarn.app.mapreduce.am.resource.mb	15702	15702
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	188416	31396
yarn.nodemanager.resource.memory-mb	188416	31396

## z1d.12xlarge

Konfigurationsoption	Standardwert	Mit installierter HBase-Anwendung
mapreduce.map.java.opts	-Xmx6417m	-Xmx6417m
mapreduce.java.opts	-Xmx12834m	-Xmx12834m
mapreduce.map.memory.mb	8021	8021
mapreduce.reduce.memory.mb	16042	16042
yarn.app.mapreduce.am.resource.mb	16042	16042
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	385024	32100
yarn.nodemanager.resource.memory-mb	385024	32100

## Hadoop-Daemon-Konfigurationseinstellungen

Hadoop-Daemon-Einstellungen unterscheiden sich je nach dem EC2-Instance-Typ, den ein Cluster-Knoten verwendet. In der folgenden Tabelle sind die Standardkonfigurationseinstellungen für die einzelnen EC2 Instance-Typen in Clustern aufgeführt.

Zum Anpassen dieser Einstellungen verwenden Sie die `hadoop-env`-Konfigurationsklassifizierung. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

### Instance-Typen

- [C1-Instances](#)
- [C3-Instances](#)
- [C4-Instances](#)
- [C5-Instances](#)

- [C5a-Instances](#)
- [C5ad-Instances](#)
- [C5d-Instances](#)
- [C5n-Instances](#)
- [C6a-Instances](#)
- [C6g-Instances](#)
- [C6gd-Instances](#)
- [C6gn-Instances](#)
- [C6i-Instances](#)
- [C6id-Instances](#)
- [C6in-Instances](#)
- [c7a-Instances](#)
- [C7g-Instances](#)
- [C7gd-Instances](#)
- [C7gn-Instances](#)
- [C7i-Instances](#)
- [D2-Instances](#)
- [D3-Instances](#)
- [D3en-Instances](#)
- [G3-Instances](#)
- [G3s-Instances](#)
- [G4dn-Instances](#)
- [G5-Instances](#)
- [H1-Instances](#)
- [I2-Instances](#)
- [I3-Instances](#)
- [I3en-Instances](#)
- [I4g-Instances](#)

- [I4i-Instances](#)
- [I4gn-Instances](#)
- [I4gen-Instances](#)
- [M1-Instances](#)
- [M2-Instances](#)
- [M3-Instances](#)
- [M4-Instances](#)
- [M5-Instances](#)
- [M5a-Instances](#)
- [M5ad-Instances](#)
- [M5d-Instances](#)
- [M5dn-Instances](#)
- [M5n-Instances](#)
- [M5zn-Instances](#)
- [M6a-Instances](#)
- [M6g-Instances](#)
- [M6gd-Instances](#)
- [M6i-Instances](#)
- [M6id-Instances](#)
- [M6idn-Instances](#)
- [M6in-Instances](#)
- [M7a-Instances](#)
- [M7g-Instances](#)
- [M7gd-Instances](#)
- [M7i-Instances](#)
- [M7i-flex-Instances](#)
- [P2-Instances](#)
- [P3-Instances](#)



- [P5-Instances](#)
- [R3-Instances](#)
- [R4-Instances](#)
- [R5-Instances](#)
- [R5a-Instances](#)
- [R5ad-Instances](#)
- [r5b-Instances](#)
- [R5d-Instances](#)
- [R5dn-Instances](#)
- [R5n-Instances](#)
- [R6a-Instances](#)
- [R6g-Instances](#)
- [R6gd-Instances](#)
- [R6i-Instances](#)
- [R6id-Instances](#)
- [R6idn-Instances](#)
- [R6in-Instances](#)
- [R7a-Instances](#)
- [R7g-Instances](#)
- [R7gd-Instances](#)
- [r7i-Instances](#)
- [R7iz-Instances](#)
- [X1-Instances](#)
- [X1e-Instances](#)
- [X2gd-Instances](#)
- [x2idn-Instances](#)
- [X2iedn-Instances](#)
- [z1d-Instances](#)

## C1-Instances

### c1.medium

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	192
YARN_PROXYSERVER_HEAPSIZE	96
YARN_NODEMANAGER_HEAPSIZE	128
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	128
HADOOP_NAMENODE_HEAPSIZE	192
HADOOP_DATANODE_HEAPSIZE	96

### c1.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	768
YARN_PROXYSERVER_HEAPSIZE	384
YARN_NODEMANAGER_HEAPSIZE	512
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	512
HADOOP_NAMENODE_HEAPSIZE	768
HADOOP_DATANODE_HEAPSIZE	384

## C3-Instances

### c3.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2124
YARN_PROXYSERVER_HEAPSIZE	2124
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2124
HADOOP_NAMENODE_HEAPSIZE	972
HADOOP_DATANODE_HEAPSIZE	588

### c3.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2396
YARN_PROXYSERVER_HEAPSIZE	2396
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2396
HADOOP_NAMENODE_HEAPSIZE	1740
HADOOP_DATANODE_HEAPSIZE	757

### c3.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2703
YARN_PROXYSERVER_HEAPSIZE	2703
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2703
HADOOP_NAMENODE_HEAPSIZE	3276
HADOOP_DATANODE_HEAPSIZE	1064

### c3.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3317
YARN_PROXYSERVER_HEAPSIZE	3317
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3317
HADOOP_NAMENODE_HEAPSIZE	6348
HADOOP_DATANODE_HEAPSIZE	1679

## C4-Instances

### c4.large

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	1152
YARN_PROXYSERVER_HEAPSIZE	1152
YARN_NODEMANAGER_HEAPSIZE	1152
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	1152
HADOOP_NAMENODE_HEAPSIZE	576
HADOOP_DATANODE_HEAPSIZE	384

### c4.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2124
YARN_PROXYSERVER_HEAPSIZE	2124
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2124
HADOOP_NAMENODE_HEAPSIZE	972
HADOOP_DATANODE_HEAPSIZE	588

## c4.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2396
YARN_PROXYSERVER_HEAPSIZE	2396
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2396
HADOOP_NAMENODE_HEAPSIZE	1740
HADOOP_DATANODE_HEAPSIZE	757

## c4.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2703
YARN_PROXYSERVER_HEAPSIZE	2703
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2703
HADOOP_NAMENODE_HEAPSIZE	3276
HADOOP_DATANODE_HEAPSIZE	1064

## c4.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3317
YARN_PROXYSERVER_HEAPSIZE	3317
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3317
HADOOP_NAMENODE_HEAPSIZE	6348
HADOOP_DATANODE_HEAPSIZE	1679

## C5-Instances

### c5.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2252
YARN_PROXYSERVER_HEAPSIZE	2252
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2252
HADOOP_NAMENODE_HEAPSIZE	1024
HADOOP_DATANODE_HEAPSIZE	614

## c5.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2416
YARN_PROXYSERVER_HEAPSIZE	2416
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2416
HADOOP_NAMENODE_HEAPSIZE	1843
HADOOP_DATANODE_HEAPSIZE	778

## c5.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2744
YARN_PROXYSERVER_HEAPSIZE	2744
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2744
HADOOP_NAMENODE_HEAPSIZE	3481
HADOOP_DATANODE_HEAPSIZE	1105



## c5.9xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3563
YARN_PROXYSERVER_HEAPSIZE	3563
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3563
HADOOP_NAMENODE_HEAPSIZE	7577
HADOOP_DATANODE_HEAPSIZE	1925

## c5.12xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4055
YARN_PROXYSERVER_HEAPSIZE	4055
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4055
HADOOP_NAMENODE_HEAPSIZE	10035
HADOOP_DATANODE_HEAPSIZE	2416

## c5.18xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	5038
YARN_PROXYSERVER_HEAPSIZE	5038
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5038
HADOOP_NAMENODE_HEAPSIZE	14950
HADOOP_DATANODE_HEAPSIZE	3399

## c5.24xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	6021
YARN_PROXYSERVER_HEAPSIZE	6021
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	6021
HADOOP_NAMENODE_HEAPSIZE	19865
HADOOP_DATANODE_HEAPSIZE	4096

## C5a-Instances

### c5a.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2124
YARN_PROXYSERVER_HEAPSIZE	2124
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2124
HADOOP_NAMENODE_HEAPSIZE	972
HADOOP_DATANODE_HEAPSIZE	588

### c5a.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

## c5a.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## c5a.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## c5a.12xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4055
YARN_PROXYSERVER_HEAPSIZE	4055
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4055
HADOOP_NAMENODE_HEAPSIZE	10035
HADOOP_DATANODE_HEAPSIZE	2416

## c5a.16xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## c5a.24xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

## C5ad-Instances

### c5ad.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2124
YARN_PROXYSERVER_HEAPSIZE	2124
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2124
HADOOP_NAMENODE_HEAPSIZE	972
HADOOP_DATANODE_HEAPSIZE	588

## c5ad.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

## c5ad.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## c5ad.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## c5ad.12xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3962
YARN_PROXYSERVER_HEAPSIZE	3962
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3962
HADOOP_NAMENODE_HEAPSIZE	9574
HADOOP_DATANODE_HEAPSIZE	2324



## c5ad.16xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## c5ad.24xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

## C5d-Instances

### c5d.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2252
YARN_PROXYSERVER_HEAPSIZE	2252
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2252
HADOOP_NAMENODE_HEAPSIZE	1024
HADOOP_DATANODE_HEAPSIZE	614

### c5d.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2416
YARN_PROXYSERVER_HEAPSIZE	2416
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2416
HADOOP_NAMENODE_HEAPSIZE	1843
HADOOP_DATANODE_HEAPSIZE	778

## c5d.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2744
YARN_PROXYSERVER_HEAPSIZE	2744
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2744
HADOOP_NAMENODE_HEAPSIZE	3481
HADOOP_DATANODE_HEAPSIZE	1105

## c5d.9xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3563
YARN_PROXYSERVER_HEAPSIZE	3563
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3563
HADOOP_NAMENODE_HEAPSIZE	7577
HADOOP_DATANODE_HEAPSIZE	1925

## c5d.12xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4055
YARN_PROXYSERVER_HEAPSIZE	4055
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4055
HADOOP_NAMENODE_HEAPSIZE	10035
HADOOP_DATANODE_HEAPSIZE	2416

## c5d.18xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	5038
YARN_PROXYSERVER_HEAPSIZE	5038
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5038
HADOOP_NAMENODE_HEAPSIZE	14950
HADOOP_DATANODE_HEAPSIZE	3399

## c5d.24xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	6021
YARN_PROXYSERVER_HEAPSIZE	6021
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	6021
HADOOP_NAMENODE_HEAPSIZE	19865
HADOOP_DATANODE_HEAPSIZE	4096

## C5n-Instances

### c5n.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2304
YARN_PROXYSERVER_HEAPSIZE	2304
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2304
HADOOP_NAMENODE_HEAPSIZE	1280
HADOOP_DATANODE_HEAPSIZE	665

## c5n.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2519
YARN_PROXYSERVER_HEAPSIZE	2519
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2519
HADOOP_NAMENODE_HEAPSIZE	2355
HADOOP_DATANODE_HEAPSIZE	880

## c5n.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2949
YARN_PROXYSERVER_HEAPSIZE	2949
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2949
HADOOP_NAMENODE_HEAPSIZE	4505
HADOOP_DATANODE_HEAPSIZE	1310

## c5n.9xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4055
YARN_PROXYSERVER_HEAPSIZE	4055
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4055
HADOOP_NAMENODE_HEAPSIZE	10035
HADOOP_DATANODE_HEAPSIZE	2416

## c5n.18xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	6021
YARN_PROXYSERVER_HEAPSIZE	6021
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	6021
HADOOP_NAMENODE_HEAPSIZE	19865
HADOOP_DATANODE_HEAPSIZE	4096

## C6a-Instances

### c6a.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2124
YARN_PROXYSERVER_HEAPSIZE	2124
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2124
HADOOP_NAMENODE_HEAPSIZE	972
HADOOP_DATANODE_HEAPSIZE	588

### c6a.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762



## c6a.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## c6a.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## c6a.12xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3962
YARN_PROXYSERVER_HEAPSIZE	3962
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3962
HADOOP_NAMENODE_HEAPSIZE	9574
HADOOP_DATANODE_HEAPSIZE	2324

## c6a.16xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## c6a.24xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

## c6a.32xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## c6a.48xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

## C6g-Instances

### c6g.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2124
YARN_PROXYSERVER_HEAPSIZE	2124
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2124
HADOOP_NAMENODE_HEAPSIZE	972
HADOOP_DATANODE_HEAPSIZE	588

## c6g.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

## c6g.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## c6g.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## c6g.12xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3962
YARN_PROXYSERVER_HEAPSIZE	3962
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3962
HADOOP_NAMENODE_HEAPSIZE	9574
HADOOP_DATANODE_HEAPSIZE	2324

## c6g.16xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## C6gd-Instances

### c6gd.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2124
YARN_PROXYSERVER_HEAPSIZE	2124
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2124
HADOOP_NAMENODE_HEAPSIZE	972
HADOOP_DATANODE_HEAPSIZE	588

## c6gd.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

## c6gd.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075



## c6gd.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## c6gd.12xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3962
YARN_PROXYSERVER_HEAPSIZE	3962
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3962
HADOOP_NAMENODE_HEAPSIZE	9574
HADOOP_DATANODE_HEAPSIZE	2324

## c6gd.16xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## C6gn-Instances

### c6gn.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2124
YARN_PROXYSERVER_HEAPSIZE	2124
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2124
HADOOP_NAMENODE_HEAPSIZE	972
HADOOP_DATANODE_HEAPSIZE	588

## c6gn.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

## c6gn.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## c6gn.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## c6gn.12xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3962
YARN_PROXYSERVER_HEAPSIZE	3962
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3962
HADOOP_NAMENODE_HEAPSIZE	9574
HADOOP_DATANODE_HEAPSIZE	2324

## c6gn.16xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## C6i-Instances

### c6i.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2124
YARN_PROXYSERVER_HEAPSIZE	2124
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2124
HADOOP_NAMENODE_HEAPSIZE	972
HADOOP_DATANODE_HEAPSIZE	588

## c6i.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

## c6i.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

**c6i.8xlarge**

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

**c6i.12xlarge**

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3962
YARN_PROXYSERVER_HEAPSIZE	3962
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3962
HADOOP_NAMENODE_HEAPSIZE	9574
HADOOP_DATANODE_HEAPSIZE	2324

## c6i.16xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## c6i.24xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096



## c6i.32xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## C6id-Instances

### c6id.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2124
YARN_PROXYSERVER_HEAPSIZE	2124
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2124
HADOOP_NAMENODE_HEAPSIZE	972
HADOOP_DATANODE_HEAPSIZE	588

## c6id.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

## c6id.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## c6id.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## c6id.12xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3962
YARN_PROXYSERVER_HEAPSIZE	3962
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3962
HADOOP_NAMENODE_HEAPSIZE	9574
HADOOP_DATANODE_HEAPSIZE	2324

## c6id.16xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## c6id.24xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

## c6id.32xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## C6in-Instances

### c6in.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2124
YARN_PROXYSERVER_HEAPSIZE	2124
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2124
HADOOP_NAMENODE_HEAPSIZE	972
HADOOP_DATANODE_HEAPSIZE	588

## c6in.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

## c6in.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## c6in.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## c6in.12xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3962
YARN_PROXYSERVER_HEAPSIZE	3962
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3962
HADOOP_NAMENODE_HEAPSIZE	9574
HADOOP_DATANODE_HEAPSIZE	2324

## c6in.16xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## c6in.24xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096



## c6in.32xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## c7a-Instances

### c7a.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2124
YARN_PROXYSERVER_HEAPSIZE	2124
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2124
HADOOP_NAMENODE_HEAPSIZE	972
HADOOP_DATANODE_HEAPSIZE	588

## c7a.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

## c7a.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## c7a.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## c7a.12xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3962
YARN_PROXYSERVER_HEAPSIZE	3962
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3962
HADOOP_NAMENODE_HEAPSIZE	9574
HADOOP_DATANODE_HEAPSIZE	2324

## c7a.16xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## c7a.24xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

## c7a.32xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## c7a.48xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

## C7g-Instances

### c7g.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2124
YARN_PROXYSERVER_HEAPSIZE	2124
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2124
HADOOP_NAMENODE_HEAPSIZE	972
HADOOP_DATANODE_HEAPSIZE	588

### c7g.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

## c7g.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## c7g.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## c7g.12xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3962
YARN_PROXYSERVER_HEAPSIZE	3962
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3962
HADOOP_NAMENODE_HEAPSIZE	9574
HADOOP_DATANODE_HEAPSIZE	2324

## c7g.16xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949



## C7gd-Instances

### c7gd.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2124
YARN_PROXYSERVER_HEAPSIZE	2124
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2124
HADOOP_NAMENODE_HEAPSIZE	972
HADOOP_DATANODE_HEAPSIZE	588

### c7gd.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

## c7gd.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## c7gd.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## c7gd.12xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3962
YARN_PROXYSERVER_HEAPSIZE	3962
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3962
HADOOP_NAMENODE_HEAPSIZE	9574
HADOOP_DATANODE_HEAPSIZE	2324

## c7gd.16xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## C7gn-Instances

### c7gn.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2124
YARN_PROXYSERVER_HEAPSIZE	2124
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2124
HADOOP_NAMENODE_HEAPSIZE	972
HADOOP_DATANODE_HEAPSIZE	588

### c7gn.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

## c7gn.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## c7gn.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## c7gn.12xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3962
YARN_PROXYSERVER_HEAPSIZE	3962
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3962
HADOOP_NAMENODE_HEAPSIZE	9574
HADOOP_DATANODE_HEAPSIZE	2324

## c7gn.16xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## C7i-Instances

### c7i.x groß

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2124
YARN_PROXYSERVER_HEAPSIZE	2124
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2124
HADOOP_NAMENODE_HEAPSIZE	972
HADOOP_DATANODE_HEAPSIZE	588

### c7i.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

## c7i.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## c7i.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699



## c7i.12xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3962
YARN_PROXYSERVER_HEAPSIZE	3962
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3962
HADOOP_NAMENODE_HEAPSIZE	9574
HADOOP_DATANODE_HEAPSIZE	2324

## c7i.16xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## c7i.24xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

## c7i.48xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

## D2-Instances

### d2.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

### d2.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## d2.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## d2.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## D3-Instances

### d3.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

### d3.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## d3.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## d3.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## D3en-Instances

### d3en.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

### d3en.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## d3en.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## d3en.6xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3962
YARN_PROXYSERVER_HEAPSIZE	3962
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3962
HADOOP_NAMENODE_HEAPSIZE	9574
HADOOP_DATANODE_HEAPSIZE	2324



## d3en.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## d3en.12xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

## G3-Instances

### g3.4xgroß

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

### g3.8xgroß

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## g3.16xgroß

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

## G3s-Instances

### g3s.xgroß

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## G4dn-Instances

### g4dn.xgroß

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2416
YARN_PROXYSERVER_HEAPSIZE	2416
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2416
HADOOP_NAMENODE_HEAPSIZE	1843
HADOOP_DATANODE_HEAPSIZE	778

### g4dn.2xgroß

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2744
YARN_PROXYSERVER_HEAPSIZE	2744
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2744
HADOOP_NAMENODE_HEAPSIZE	3481
HADOOP_DATANODE_HEAPSIZE	1105

## g4dn.4xgroß

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3399
YARN_PROXYSERVER_HEAPSIZE	3399
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3399
HADOOP_NAMENODE_HEAPSIZE	6758
HADOOP_DATANODE_HEAPSIZE	1761

## g4dn.8xgroß

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4710
YARN_PROXYSERVER_HEAPSIZE	4710
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4710
HADOOP_NAMENODE_HEAPSIZE	13312
HADOOP_DATANODE_HEAPSIZE	3072

## g4dn.12xgroß

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	6021
YARN_PROXYSERVER_HEAPSIZE	6021
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	6021
HADOOP_NAMENODE_HEAPSIZE	19865
HADOOP_DATANODE_HEAPSIZE	4096

## g4dn.16xgroß

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	7331
YARN_PROXYSERVER_HEAPSIZE	7331
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7331
HADOOP_NAMENODE_HEAPSIZE	26419
HADOOP_DATANODE_HEAPSIZE	4096

## G5-Instances

### g5.xgroß

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

### g5.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## g5.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## g5.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949



## g5.12xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

## g5.16xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## g5.24xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

## g5.48xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

## H1-Instances

### h1.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2744
YARN_PROXYSERVER_HEAPSIZE	2744
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2744
HADOOP_NAMENODE_HEAPSIZE	3481
HADOOP_DATANODE_HEAPSIZE	1105

### h1.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3399
YARN_PROXYSERVER_HEAPSIZE	3399
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3399
HADOOP_NAMENODE_HEAPSIZE	6758
HADOOP_DATANODE_HEAPSIZE	1761

## h1.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4710
YARN_PROXYSERVER_HEAPSIZE	4710
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4710
HADOOP_NAMENODE_HEAPSIZE	13312
HADOOP_DATANODE_HEAPSIZE	3072

## h1.16xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	7331
YARN_PROXYSERVER_HEAPSIZE	7331
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7331
HADOOP_NAMENODE_HEAPSIZE	26419
HADOOP_DATANODE_HEAPSIZE	4096

## I2-Instances

### i2.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

### i2.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## i2.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## i2.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## I3-Instances

### i3.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

### i3.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## i3.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## i3.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096



### i3.16xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

### I3en-Instances

#### i3en.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2744
YARN_PROXYSERVER_HEAPSIZE	2744
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2744
HADOOP_NAMENODE_HEAPSIZE	3481
HADOOP_DATANODE_HEAPSIZE	1105

## i3en.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3399
YARN_PROXYSERVER_HEAPSIZE	3399
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3399
HADOOP_NAMENODE_HEAPSIZE	6758
HADOOP_DATANODE_HEAPSIZE	1761

## i3en.3xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4055
YARN_PROXYSERVER_HEAPSIZE	4055
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4055
HADOOP_NAMENODE_HEAPSIZE	10035
HADOOP_DATANODE_HEAPSIZE	2416

## i3en.6xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	6021
YARN_PROXYSERVER_HEAPSIZE	6021
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	6021
HADOOP_NAMENODE_HEAPSIZE	19865
HADOOP_DATANODE_HEAPSIZE	4096

## i3en.12xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	9953
YARN_PROXYSERVER_HEAPSIZE	9953
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9953
HADOOP_NAMENODE_HEAPSIZE	39526
HADOOP_DATANODE_HEAPSIZE	4096

## i3en.24xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	17817
YARN_PROXYSERVER_HEAPSIZE	17817
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	17817
HADOOP_NAMENODE_HEAPSIZE	78848
HADOOP_DATANODE_HEAPSIZE	4096

## I4g-Instances

### i4g.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## i4g.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## i4g.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## i4g.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## i4g.16xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

## I4i-Instances

### i4i.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

### i4i.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## i4i.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## i4i.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096



## i4i.12xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

## i4i.16xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

## i4i.24xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

## i4i.32xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	22077
YARN_PROXYSERVER_HEAPSIZE	22077
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	22077
HADOOP_NAMENODE_HEAPSIZE	100147
HADOOP_DATANODE_HEAPSIZE	4096

## Im4gn-Instances

### im4gn.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

### im4gn.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## im4gn.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## im4gn.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## im4gn.16xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## Is4gen-Instances

### is4gen.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2557
YARN_PROXYSERVER_HEAPSIZE	2557
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2557
HADOOP_NAMENODE_HEAPSIZE	2547
HADOOP_DATANODE_HEAPSIZE	919

## is4gen.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3025
YARN_PROXYSERVER_HEAPSIZE	3025
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3025
HADOOP_NAMENODE_HEAPSIZE	4889
HADOOP_DATANODE_HEAPSIZE	1387

## is4gen.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3962
YARN_PROXYSERVER_HEAPSIZE	3962
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3962
HADOOP_NAMENODE_HEAPSIZE	9574
HADOOP_DATANODE_HEAPSIZE	2324

## is4gen.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

## M1-Instances

### m1.small

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	256
YARN_PROXYSERVER_HEAPSIZE	96
YARN_NODEMANAGER_HEAPSIZE	192
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	128
HADOOP_NAMENODE_HEAPSIZE	192
HADOOP_DATANODE_HEAPSIZE	96

## m1.medium

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	384
YARN_PROXYSERVER_HEAPSIZE	192
YARN_NODEMANAGER_HEAPSIZE	256
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	256
HADOOP_NAMENODE_HEAPSIZE	384
HADOOP_DATANODE_HEAPSIZE	192

## m1.large

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	768
YARN_PROXYSERVER_HEAPSIZE	384
YARN_NODEMANAGER_HEAPSIZE	512
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	512
HADOOP_NAMENODE_HEAPSIZE	768
HADOOP_DATANODE_HEAPSIZE	384



## m1.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	1024
YARN_PROXYSERVER_HEAPSIZE	512
YARN_NODEMANAGER_HEAPSIZE	768
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	1024
HADOOP_NAMENODE_HEAPSIZE	2304
HADOOP_DATANODE_HEAPSIZE	384

## M2-Instances

### m2.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	1536
YARN_PROXYSERVER_HEAPSIZE	1024
YARN_NODEMANAGER_HEAPSIZE	1024
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	1024
HADOOP_NAMENODE_HEAPSIZE	3072
HADOOP_DATANODE_HEAPSIZE	384

## m2.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	1536
YARN_PROXYSERVER_HEAPSIZE	1024
YARN_NODEMANAGER_HEAPSIZE	1024
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	1536
HADOOP_NAMENODE_HEAPSIZE	6144
HADOOP_DATANODE_HEAPSIZE	384

## m2.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2048
YARN_PROXYSERVER_HEAPSIZE	1024
YARN_NODEMANAGER_HEAPSIZE	1536
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	1536
HADOOP_NAMENODE_HEAPSIZE	12288
HADOOP_DATANODE_HEAPSIZE	384

## M3-Instances

### m3.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2396
YARN_PROXYSERVER_HEAPSIZE	2396
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2396
HADOOP_NAMENODE_HEAPSIZE	1740
HADOOP_DATANODE_HEAPSIZE	757

### m3.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2703
YARN_PROXYSERVER_HEAPSIZE	2703
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2703
HADOOP_NAMENODE_HEAPSIZE	3276
HADOOP_DATANODE_HEAPSIZE	1064

## M4-Instances

### m4.large

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2252
YARN_PROXYSERVER_HEAPSIZE	2252
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2252
HADOOP_NAMENODE_HEAPSIZE	1024
HADOOP_DATANODE_HEAPSIZE	614

### m4.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2416
YARN_PROXYSERVER_HEAPSIZE	2416
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2416
HADOOP_NAMENODE_HEAPSIZE	1843
HADOOP_DATANODE_HEAPSIZE	778

## m4.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2744
YARN_PROXYSERVER_HEAPSIZE	2744
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2744
HADOOP_NAMENODE_HEAPSIZE	3481
HADOOP_DATANODE_HEAPSIZE	1105

## m4.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3399
YARN_PROXYSERVER_HEAPSIZE	3399
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3399
HADOOP_NAMENODE_HEAPSIZE	6758
HADOOP_DATANODE_HEAPSIZE	1761

## m4.10xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	5365
YARN_PROXYSERVER_HEAPSIZE	5365
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5365
HADOOP_NAMENODE_HEAPSIZE	16588
HADOOP_DATANODE_HEAPSIZE	3727

## m4.16xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	7331
YARN_PROXYSERVER_HEAPSIZE	7331
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7331
HADOOP_NAMENODE_HEAPSIZE	26419
HADOOP_DATANODE_HEAPSIZE	4096

## M5-Instances

### m5.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2416
YARN_PROXYSERVER_HEAPSIZE	2416
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2416
HADOOP_NAMENODE_HEAPSIZE	1843
HADOOP_DATANODE_HEAPSIZE	778

### m5.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2744
YARN_PROXYSERVER_HEAPSIZE	2744
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2744
HADOOP_NAMENODE_HEAPSIZE	3481
HADOOP_DATANODE_HEAPSIZE	1105

## m5.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3399
YARN_PROXYSERVER_HEAPSIZE	3399
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3399
HADOOP_NAMENODE_HEAPSIZE	6758
HADOOP_DATANODE_HEAPSIZE	1761

## m5.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4710
YARN_PROXYSERVER_HEAPSIZE	4710
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4710
HADOOP_NAMENODE_HEAPSIZE	13312
HADOOP_DATANODE_HEAPSIZE	3072



## m5.12xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	6021
YARN_PROXYSERVER_HEAPSIZE	6021
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	6021
HADOOP_NAMENODE_HEAPSIZE	19865
HADOOP_DATANODE_HEAPSIZE	4096

## m5.16xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	7331
YARN_PROXYSERVER_HEAPSIZE	7331
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7331
HADOOP_NAMENODE_HEAPSIZE	26419
HADOOP_DATANODE_HEAPSIZE	4096

## m5.24xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	9953
YARN_PROXYSERVER_HEAPSIZE	9953
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9953
HADOOP_NAMENODE_HEAPSIZE	39526
HADOOP_DATANODE_HEAPSIZE	4096

## M5a-Instances

### m5a.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2416
YARN_PROXYSERVER_HEAPSIZE	2416
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2416
HADOOP_NAMENODE_HEAPSIZE	1843
HADOOP_DATANODE_HEAPSIZE	778

## m5a.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2744
YARN_PROXYSERVER_HEAPSIZE	2744
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2744
HADOOP_NAMENODE_HEAPSIZE	3481
HADOOP_DATANODE_HEAPSIZE	1105

## m5a.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3399
YARN_PROXYSERVER_HEAPSIZE	3399
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3399
HADOOP_NAMENODE_HEAPSIZE	6758
HADOOP_DATANODE_HEAPSIZE	1761

## m5a.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4710
YARN_PROXYSERVER_HEAPSIZE	4710
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4710
HADOOP_NAMENODE_HEAPSIZE	13312
HADOOP_DATANODE_HEAPSIZE	3072

## m5a.12xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	6021
YARN_PROXYSERVER_HEAPSIZE	6021
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	6021
HADOOP_NAMENODE_HEAPSIZE	19865
HADOOP_DATANODE_HEAPSIZE	4096

## m5a.16xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	7331
YARN_PROXYSERVER_HEAPSIZE	7331
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7331
HADOOP_NAMENODE_HEAPSIZE	26419
HADOOP_DATANODE_HEAPSIZE	4096

## m5a.24xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	9953
YARN_PROXYSERVER_HEAPSIZE	9953
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9953
HADOOP_NAMENODE_HEAPSIZE	39526
HADOOP_DATANODE_HEAPSIZE	4096

## M5ad-Instances

### m5ad.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

### m5ad.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## m5ad.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## m5ad.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## m5ad.12xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

## m5ad.16xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096



## m5ad.24xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

## M5d-Instances

### m5d.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2416
YARN_PROXYSERVER_HEAPSIZE	2416
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2416
HADOOP_NAMENODE_HEAPSIZE	1843
HADOOP_DATANODE_HEAPSIZE	778

## m5d.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2744
YARN_PROXYSERVER_HEAPSIZE	2744
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2744
HADOOP_NAMENODE_HEAPSIZE	3481
HADOOP_DATANODE_HEAPSIZE	1105

## m5d.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3399
YARN_PROXYSERVER_HEAPSIZE	3399
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3399
HADOOP_NAMENODE_HEAPSIZE	6758
HADOOP_DATANODE_HEAPSIZE	1761

## m5d.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4710
YARN_PROXYSERVER_HEAPSIZE	4710
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4710
HADOOP_NAMENODE_HEAPSIZE	13312
HADOOP_DATANODE_HEAPSIZE	3072

## m5d.12xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	6021
YARN_PROXYSERVER_HEAPSIZE	6021
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	6021
HADOOP_NAMENODE_HEAPSIZE	19865
HADOOP_DATANODE_HEAPSIZE	4096

## m5d.16xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	7331
YARN_PROXYSERVER_HEAPSIZE	7331
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7331
HADOOP_NAMENODE_HEAPSIZE	26419
HADOOP_DATANODE_HEAPSIZE	4096

## m5d.24xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	9953
YARN_PROXYSERVER_HEAPSIZE	9953
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9953
HADOOP_NAMENODE_HEAPSIZE	39526
HADOOP_DATANODE_HEAPSIZE	4096

## M5dn-Instances

### m5dn.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

### m5dn.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## m5dn.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## m5dn.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## m5dn.12xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

## m5dn.16xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## m5dn.24xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

## M5n-Instances

### m5n.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762



## m5n.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## m5n.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## m5n.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## m5n.12xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

## m5n.16xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## m5n.24xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

## M5zn-Instances

### m5zn.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2396
YARN_PROXYSERVER_HEAPSIZE	2396
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2396
HADOOP_NAMENODE_HEAPSIZE	1740
HADOOP_DATANODE_HEAPSIZE	757

### m5zn.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## m5zn.3xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3025
YARN_PROXYSERVER_HEAPSIZE	3025
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3025
HADOOP_NAMENODE_HEAPSIZE	4889
HADOOP_DATANODE_HEAPSIZE	1387

## m5zn.6xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3962
YARN_PROXYSERVER_HEAPSIZE	3962
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3962
HADOOP_NAMENODE_HEAPSIZE	9574
HADOOP_DATANODE_HEAPSIZE	2324

## m5zn.12xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

## M6a-Instances

### m6a.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

## m6a.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## m6a.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## m6a.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## m6a.12xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096



## m6a.16xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## m6a.24xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

## m6a.32xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

## m6a.48xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

## M6g-Instances

### m6g.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

### m6g.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## m6g.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## m6g.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## m6g.12xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	5877
YARN_PROXYSERVER_HEAPSIZE	5877
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5877
HADOOP_NAMENODE_HEAPSIZE	19148
HADOOP_DATANODE_HEAPSIZE	4096

## m6g.16xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## M6gd-Instances

### m6gd.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

### m6gd.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## m6gd.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## m6gd.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## m6gd.12xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	5877
YARN_PROXYSERVER_HEAPSIZE	5877
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5877
HADOOP_NAMENODE_HEAPSIZE	19148
HADOOP_DATANODE_HEAPSIZE	4096

## m6gd.16xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096



## M6i-Instances

### m6i.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

### m6i.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## m6i.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## m6i.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## m6i.12xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	5877
YARN_PROXYSERVER_HEAPSIZE	5877
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5877
HADOOP_NAMENODE_HEAPSIZE	19148
HADOOP_DATANODE_HEAPSIZE	4096

## m6i.16xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## m6i.24xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

## m6i.32xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

## M6id-Instances

### m6id.x groß

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

### m6id.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## m6id.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## m6id.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## m6id.12xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

## m6id.16xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## m6id.24xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

## m6id.32xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096



## M6idn-Instances

### m6idn.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

### m6id.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## m6idn.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## m6idn.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## m6idn.12xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

## m6idn.16xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## m6idn.24xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

## m6idn.32xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

## M6in-Instances

### m6in.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

### m6in.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## m6in.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## m6n.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## m6in.12xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	5877
YARN_PROXYSERVER_HEAPSIZE	5877
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5877
HADOOP_NAMENODE_HEAPSIZE	19148
HADOOP_DATANODE_HEAPSIZE	4096

## m6in.16xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## m6in.24xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

## m6in.32xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096



## M7a-Instances

### m7a.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

### m7a.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## m7a.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## m7a.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## m7a.12xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

## m7a.16xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## m7a.24xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

## m7a.32xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

## m7a.48xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

## M7g-Instances

### m7g.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

## m7g.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## m7g.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## m7g.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## m7g.12xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

## m7g.16xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## M7gd-Instances

### m7gd.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762



## m7gd.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## m7gd.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## m7gd.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## m7gd.12xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

## m7gd.16xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## M7i-Instances

### m7i.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

## m7i.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## m7i.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## m7i.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## m7i.12xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	5877
YARN_PROXYSERVER_HEAPSIZE	5877
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5877
HADOOP_NAMENODE_HEAPSIZE	19148
HADOOP_DATANODE_HEAPSIZE	4096

## m7i.16xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## m7i.24xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

## m7i.48xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

## M7i-flex-Instances

### m7i-flex.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

## m7i-flex.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## m7i-flex.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699



## m7i-flex.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## P2-Instances

### p2.xgroß

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## p2.8xgroß

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

## p2.16xgroß

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

## P3-Instances

### p3.2xgroß

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

### p3.8xgroß

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## p3.16xgroß

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

## P5-Instances

### p5.48xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	42065
YARN_PROXYSERVER_HEAPSIZE	42065
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	42065
HADOOP_NAMENODE_HEAPSIZE	200089
HADOOP_DATANODE_HEAPSIZE	4096

## R3-Instances

### r3.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

### r3.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## r3.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## r3.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## R4-Instances

### r4.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

### r4.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## r4.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## r4.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096



## r4.16xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

## R5-Instances

### r5.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2744
YARN_PROXYSERVER_HEAPSIZE	2744
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2744
HADOOP_NAMENODE_HEAPSIZE	3481
HADOOP_DATANODE_HEAPSIZE	1105

## r5.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3399
YARN_PROXYSERVER_HEAPSIZE	3399
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3399
HADOOP_NAMENODE_HEAPSIZE	6758
HADOOP_DATANODE_HEAPSIZE	1761

## r5.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4710
YARN_PROXYSERVER_HEAPSIZE	4710
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4710
HADOOP_NAMENODE_HEAPSIZE	13312
HADOOP_DATANODE_HEAPSIZE	3072

## r5.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	7331
YARN_PROXYSERVER_HEAPSIZE	7331
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7331
HADOOP_NAMENODE_HEAPSIZE	26419
HADOOP_DATANODE_HEAPSIZE	4096

## r5.12xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	9953
YARN_PROXYSERVER_HEAPSIZE	9953
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9953
HADOOP_NAMENODE_HEAPSIZE	39526
HADOOP_DATANODE_HEAPSIZE	4096

## r5.16xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	12574
YARN_PROXYSERVER_HEAPSIZE	12574
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12574
HADOOP_NAMENODE_HEAPSIZE	52633
HADOOP_DATANODE_HEAPSIZE	4096

## r5.24xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	17817
YARN_PROXYSERVER_HEAPSIZE	17817
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	17817
HADOOP_NAMENODE_HEAPSIZE	78848
HADOOP_DATANODE_HEAPSIZE	4096

## R5a-Instances

### r5a.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2744
YARN_PROXYSERVER_HEAPSIZE	2744
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2744
HADOOP_NAMENODE_HEAPSIZE	3481
HADOOP_DATANODE_HEAPSIZE	1105

### r5a.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3399
YARN_PROXYSERVER_HEAPSIZE	3399
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3399
HADOOP_NAMENODE_HEAPSIZE	6758
HADOOP_DATANODE_HEAPSIZE	1761

## r5a.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4710
YARN_PROXYSERVER_HEAPSIZE	4710
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4710
HADOOP_NAMENODE_HEAPSIZE	13312
HADOOP_DATANODE_HEAPSIZE	3072

## r5a.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	7331
YARN_PROXYSERVER_HEAPSIZE	7331
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7331
HADOOP_NAMENODE_HEAPSIZE	26419
HADOOP_DATANODE_HEAPSIZE	4096

## r5a.12xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	9953
YARN_PROXYSERVER_HEAPSIZE	9953
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9953
HADOOP_NAMENODE_HEAPSIZE	39526
HADOOP_DATANODE_HEAPSIZE	4096

## r5a.16xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	12574
YARN_PROXYSERVER_HEAPSIZE	12574
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12574
HADOOP_NAMENODE_HEAPSIZE	52633
HADOOP_DATANODE_HEAPSIZE	4096

## r5a.24xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	17817
YARN_PROXYSERVER_HEAPSIZE	17817
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	17817
HADOOP_NAMENODE_HEAPSIZE	78848
HADOOP_DATANODE_HEAPSIZE	4096

## R5ad-Instances

### r5ad.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075



## r5ad.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## r5ad.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## r5ad.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## r5ad.12xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

## r5ad.16xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	12247
YARN_PROXYSERVER_HEAPSIZE	12247
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12247
HADOOP_NAMENODE_HEAPSIZE	50995
HADOOP_DATANODE_HEAPSIZE	4096

## r5ad.24xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

## r5b-Instances

### r5b.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

### r5b.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## r5b.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## r5b.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## r5b.12xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

## r5b.16xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

## r5b.24xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

## R5d-Instances

### r5d.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2744
YARN_PROXYSERVER_HEAPSIZE	2744
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2744
HADOOP_NAMENODE_HEAPSIZE	3481
HADOOP_DATANODE_HEAPSIZE	1105

## r5d.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3399
YARN_PROXYSERVER_HEAPSIZE	3399
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3399
HADOOP_NAMENODE_HEAPSIZE	6758
HADOOP_DATANODE_HEAPSIZE	1761

## r5d.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4710
YARN_PROXYSERVER_HEAPSIZE	4710
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4710
HADOOP_NAMENODE_HEAPSIZE	13312
HADOOP_DATANODE_HEAPSIZE	3072



## r5d.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	7331
YARN_PROXYSERVER_HEAPSIZE	7331
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7331
HADOOP_NAMENODE_HEAPSIZE	26419
HADOOP_DATANODE_HEAPSIZE	4096

## r5d.12xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	9953
YARN_PROXYSERVER_HEAPSIZE	9953
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9953
HADOOP_NAMENODE_HEAPSIZE	39526
HADOOP_DATANODE_HEAPSIZE	4096

## r5d.16xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	12574
YARN_PROXYSERVER_HEAPSIZE	12574
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12574
HADOOP_NAMENODE_HEAPSIZE	52633
HADOOP_DATANODE_HEAPSIZE	4096

## r5d.24xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	17817
YARN_PROXYSERVER_HEAPSIZE	17817
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	17817
HADOOP_NAMENODE_HEAPSIZE	78848
HADOOP_DATANODE_HEAPSIZE	4096

## R5dn-Instances

### r5dn.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

### r5dn.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## r5dn.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## r5dn.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## r5dn.12xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

## r5dn.16xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

## r5dn.24xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

## R5n-Instances

### r5n.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## r5n.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## r5n.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## r5n.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## r5n.12xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096



## r5n.16xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

## r5n.24xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

## R6a-Instances

### r6a.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

### r6a.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## r6a.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## r6a.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## r6a.12xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

## r6a.16xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

## r6a.24xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

## r6a.32xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	22077
YARN_PROXYSERVER_HEAPSIZE	22077
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	22077
HADOOP_NAMENODE_HEAPSIZE	100147
HADOOP_DATANODE_HEAPSIZE	4096

## r6a.48xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	32071
YARN_PROXYSERVER_HEAPSIZE	32071
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	32071
HADOOP_NAMENODE_HEAPSIZE	150118
HADOOP_DATANODE_HEAPSIZE	4096

## R6g-Instances

### r6g.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## r6g.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## r6g.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## r6g.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## r6g.12xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096



## r6g.16xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

## R6gd-Instances

### r6gd.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## r6gd.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## r6gd.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## r6gd.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## r6gd.12xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

## r6gd.16xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

## R6i-Instances

### r6i.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## r6i.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## r6i.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## r6i.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## r6i.12xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

## r6i.16xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

## r6i.24xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

## r6i.32xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	21544
YARN_PROXYSERVER_HEAPSIZE	21544
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	21544
HADOOP_NAMENODE_HEAPSIZE	97484
HADOOP_DATANODE_HEAPSIZE	4096

## R6id-Instances

### r6id.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075



## r6id.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## r6id.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## r6id.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## r6id.12xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

## r6id.16xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

## r6id.24xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

## r6id.32xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	22077
YARN_PROXYSERVER_HEAPSIZE	22077
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	22077
HADOOP_NAMENODE_HEAPSIZE	100147
HADOOP_DATANODE_HEAPSIZE	4096

## R6idn-Instances

## r6idn.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## r6idn.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## r6idn.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## r6idn.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## r6idn.12xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

## r6idn.16xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

## r6idn.24xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

## r6idn.32xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	22077
YARN_PROXYSERVER_HEAPSIZE	22077
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	22077
HADOOP_NAMENODE_HEAPSIZE	100147
HADOOP_DATANODE_HEAPSIZE	4096

## R6in-Instances

### r6in.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075



## r6in.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## r6in.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## r6in.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## r6in.12xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

## r6in.16xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

## r6in.24xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

## r6in.32xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	22077
YARN_PROXYSERVER_HEAPSIZE	22077
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	22077
HADOOP_NAMENODE_HEAPSIZE	100147
HADOOP_DATANODE_HEAPSIZE	4096

## R7a-Instances

## r7a.x groß

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## r7a.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## r7a.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## r7a.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## r7a.12xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

## r7a.16xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

## r7a.24xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

## r7a.32xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	22077
YARN_PROXYSERVER_HEAPSIZE	22077
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	22077
HADOOP_NAMENODE_HEAPSIZE	100147
HADOOP_DATANODE_HEAPSIZE	4096

## r7a.48xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	32071
YARN_PROXYSERVER_HEAPSIZE	32071
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	32071
HADOOP_NAMENODE_HEAPSIZE	150118
HADOOP_DATANODE_HEAPSIZE	4096



## R7g-Instances

### r7g.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

### r7g.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## r7g.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## r7g.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## r7g.12xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

## r7g.16xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

## R7gd-Instances

### r7gd.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

### r7gd.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## r7gd.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## r7gd.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## r7gd.12xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

## r7gd.16xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

## r7i-Instances

### r7i.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

### r7i.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## r7i.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## r7i.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096



## r7i.12xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

## r7i.16xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

## r7i.24xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

## r7i.48xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	32071
YARN_PROXYSERVER_HEAPSIZE	32071
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	32071
HADOOP_NAMENODE_HEAPSIZE	150118
HADOOP_DATANODE_HEAPSIZE	4096

## R7iz-Instances

### r7iz.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

### r7iz.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## r7iz.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## r7iz.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## r7iz.12xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

## r7iz.16xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

## r7iz.32xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	21544
YARN_PROXYSERVER_HEAPSIZE	21544
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	21544
HADOOP_NAMENODE_HEAPSIZE	97484
HADOOP_DATANODE_HEAPSIZE	4096

## X1-Instances

## x1.16xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	21544
YARN_PROXYSERVER_HEAPSIZE	21544
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	21544
HADOOP_NAMENODE_HEAPSIZE	97484
HADOOP_DATANODE_HEAPSIZE	4096

## x1.32xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	41000
YARN_PROXYSERVER_HEAPSIZE	41000
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	41000
HADOOP_NAMENODE_HEAPSIZE	194764
HADOOP_DATANODE_HEAPSIZE	4096

## X1e-Instances

### x1e.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4520
YARN_PROXYSERVER_HEAPSIZE	4520
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4520
HADOOP_NAMENODE_HEAPSIZE	12364
HADOOP_DATANODE_HEAPSIZE	2882

## x1e.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	6952
YARN_PROXYSERVER_HEAPSIZE	6952
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	6952
HADOOP_NAMENODE_HEAPSIZE	24524
HADOOP_DATANODE_HEAPSIZE	4096

## x1e.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	11816
YARN_PROXYSERVER_HEAPSIZE	11816
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	11816
HADOOP_NAMENODE_HEAPSIZE	48844
HADOOP_DATANODE_HEAPSIZE	4096



## x1e.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	21544
YARN_PROXYSERVER_HEAPSIZE	21544
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	21544
HADOOP_NAMENODE_HEAPSIZE	97484
HADOOP_DATANODE_HEAPSIZE	4096

## x1e.16xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	41000
YARN_PROXYSERVER_HEAPSIZE	41000
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	41000
HADOOP_NAMENODE_HEAPSIZE	194764
HADOOP_DATANODE_HEAPSIZE	4096

## x1e.32xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	79912
YARN_PROXYSERVER_HEAPSIZE	79912
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	79912
HADOOP_NAMENODE_HEAPSIZE	389324
HADOOP_DATANODE_HEAPSIZE	4096

## X2gd-Instances

### x2gd.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## x2gd.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## x2gd.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## x2gd.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

## x2gd.12xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

## x2gd.16xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	22077
YARN_PROXYSERVER_HEAPSIZE	22077
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	22077
HADOOP_NAMENODE_HEAPSIZE	100147
HADOOP_DATANODE_HEAPSIZE	4096

## x2idn-Instances

### x2idn.16xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	22077
YARN_PROXYSERVER_HEAPSIZE	22077
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	22077
HADOOP_NAMENODE_HEAPSIZE	100147
HADOOP_DATANODE_HEAPSIZE	4096

## x2idn.24xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	32071
YARN_PROXYSERVER_HEAPSIZE	32071
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	32071
HADOOP_NAMENODE_HEAPSIZE	150118
HADOOP_DATANODE_HEAPSIZE	4096

## x2idn.32xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	42065
YARN_PROXYSERVER_HEAPSIZE	42065
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	42065
HADOOP_NAMENODE_HEAPSIZE	200089
HADOOP_DATANODE_HEAPSIZE	4096

## X2iedn-Instances

### x2iedn.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

### x2iedn.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## x2iedn.4xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

## x2iedn.8xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	22077
YARN_PROXYSERVER_HEAPSIZE	22077
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	22077
HADOOP_NAMENODE_HEAPSIZE	100147
HADOOP_DATANODE_HEAPSIZE	4096



## x2iedn.16xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	42065
YARN_PROXYSERVER_HEAPSIZE	42065
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	42065
HADOOP_NAMENODE_HEAPSIZE	200089
HADOOP_DATANODE_HEAPSIZE	4096

## x2iedn.24xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	62054
YARN_PROXYSERVER_HEAPSIZE	62054
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	62054
HADOOP_NAMENODE_HEAPSIZE	300032
HADOOP_DATANODE_HEAPSIZE	4096

## x2iedn.32xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	82042
YARN_PROXYSERVER_HEAPSIZE	82042
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	82042
HADOOP_NAMENODE_HEAPSIZE	399974
HADOOP_DATANODE_HEAPSIZE	4096

## z1d-Instances

### z1d.xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	2744
YARN_PROXYSERVER_HEAPSIZE	2744
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2744
HADOOP_NAMENODE_HEAPSIZE	3481
HADOOP_DATANODE_HEAPSIZE	1105

## z1d.2xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	3399
YARN_PROXYSERVER_HEAPSIZE	3399
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3399
HADOOP_NAMENODE_HEAPSIZE	6758
HADOOP_DATANODE_HEAPSIZE	1761

## z1d.3xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	4055
YARN_PROXYSERVER_HEAPSIZE	4055
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4055
HADOOP_NAMENODE_HEAPSIZE	10035
HADOOP_DATANODE_HEAPSIZE	2416

## z1d.6xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	6021
YARN_PROXYSERVER_HEAPSIZE	6021
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	6021
HADOOP_NAMENODE_HEAPSIZE	19865
HADOOP_DATANODE_HEAPSIZE	4096

## z1d.12xlarge

Parameter	Wert
YARN_RESOURCEMANAGER_HEAPSIZE	9953
YARN_PROXYSERVER_HEAPSIZE	9953
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9953
HADOOP_NAMENODE_HEAPSIZE	39526
HADOOP_DATANODE_HEAPSIZE	4096

## HDFS-Konfiguration

Die folgende Tabelle beschreibt die standardmäßigen Hadoop Distributed File System (HDFS)-Parameter und ihre Einstellungen. Sie können diese Werte mithilfe der `hdfs-site-`

Konfigurationsklassifizierung ändern. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

### Warning

1. Das Festlegen von `dfs.replication` auf 1 auf Clustern mit weniger als vier Knoten kann zu einem HDFS-Datenverlust führen, wenn ein einzelner Knoten ausfällt. Wenn Ihr Cluster über HDFS-Speicher verfügt, empfehlen wir, den Cluster mit mindestens vier Core-Knoten für Produktionsworkloads zu konfigurieren, um Datenverlust zu vermeiden.
2. Amazon EMR erlaubt Clustern nicht, Core-Knoten unter `dfs.replication` zu skalieren. Bei `dfs.replication = 2` z. B. beträgt die Mindestanzahl von Core-Knoten 2.
3. Wenn Sie Verwaltete Skalierung oder Auto-Scaling verwenden oder die Größe Ihres Clusters manuell ändern möchten, empfehlen wir Ihnen, die Größe `dfs.replication` auf 2 oder höher einzustellen.

Parameter	Definition	Standardwert
<code>dfs.block.size</code>	Dieser Parameter betrifft die Größe der HDFS-Blöcke. Wenn er auf Daten in HDFS angewendet wird, entspricht die Aufteilungsgröße in der Regel der Größe eines HDFS-Blocks. Größere Größen bieten weniger Aufgabengranularität, belasten den Cluster NameNode jedoch auch weniger.	134217728 (128 MB)
<code>dfs.replication</code>	Die Anzahl der Kopien der einzelnen Blöcke, die dauerhaft gespeichert werden sollen. Amazon EMR legt diesen Wert auf der Grundlage der Anzahl der Core-Knoten fest, mit denen der Cluster bereitgestellt wird. Passen Sie den Wert entsprechend Ihren Anforderungen an. Zum Überschreiben des Standardwerts verwenden Sie die <code>hdfs-site</code> -Klassifizierung.	1 für Cluster, die mit weniger als vier Core-Knoten bereitgestellt werden  2 für Cluster, die mit weniger als zehn Core-Knoten bereitgestellt werden

Parameter	Definition	Standardwert
		3 für alle anderen Cluster

## Transparente Verschlüsselung in HDFS auf Amazon EMR

Die transparente Verschlüsselung wird durch die Verwendung von HDFS-Verschlüsselungszonen implementiert, wobei es sich dabei um HDFS-Pfade handelt, die von Ihnen definiert werden. Jede Verschlüsselungszone verfügt über ihren eigenen Schlüssel, der unter Verwendung der `hdfs-site`-Konfigurationsklassifikation auf dem Schlüssel-Server gespeichert wird.

Ab Amazon-EMR-Version 4.8.0 können Sie anhand der Amazon-EMR-Sicherheitskonfigurationen Datenverschlüsselungseinstellungen für Cluster einfacher konfigurieren. Sicherheitskonfigurationen bieten Einstellungen, um die Sicherheit für Daten während der Übertragung und für Daten im Ruhezustand in Amazon Elastic Block Store (Amazon EBS)-Speichervolumen sowie EMRFS-Daten in Amazon S3 zu unterstützen. Weitere Informationen finden Sie unter [Daten während der Übertragung und im Ruhezustand verschlüsseln](#) im Verwaltungshandbuch für Amazon EMR.

Amazon EMR verwendet standardmäßig den Hadoop KMS. Sie können jedoch einen anderen KMS verwenden, der den KeyProvider API-Vorgang implementiert. Jede Datei in einer HDFS-Zone verfügt über ihren eigenen, eindeutigen Datenverschlüsselungsschlüssel, der durch den Verschlüsselungszonenschlüssel verschlüsselt ist. HDFS-Daten werden verschlüsselt end-to-end (im Ruhezustand und während der Übertragung), wenn Daten in eine Verschlüsselungszone geschrieben werden, da Verschlüsselungs- und Entschlüsselungsaktivitäten nur im Client stattfinden.

Sie können keine Dateien zwischen den Verschlüsselungszonen oder von einer Verschlüsselungszone nach unverschlüsselten Pfaden verschieben.

Der HDFS-Client NameNode und der HDFS-Client interagieren über den API-Vorgang mit dem Hadoop KMS (oder einem von Ihnen konfigurierten alternativen KMS). KeyProvider Das KMS speichert Verschlüsselungsschlüssel im Unterstützungsschlüsselspeicher. Außerdem enthält Amazon EMR die JCE-Richtlinie für unbegrenzte Stärke, so dass Sie Schlüssel mit einer gewünschten Länge erstellen können.

Weitere Informationen finden Sie unter [Transparente Verschlüsselung in HDFS](#) in der Hadoop-Dokumentation.

**Note**

In Amazon EMR ist KMS über HTTPS nicht standardmäßig im Hadoop-KMS aktiviert. Weitere Informationen zum Aktivieren von KMS über HTTPS finden Sie in der [Hadoop-KMS-Dokumentation](#).

## Konfigurieren der transparenten HDFS-Verschlüsselung

Sie können die transparente Verschlüsselung in Amazon EMR konfigurieren, indem Sie Schlüssel erstellen und Verschlüsselungszonen hinzufügen. Hierfür gibt es mehrere Möglichkeiten:

- Verwenden der Amazon-EMR-Konfigurations-API-Operation beim Erstellen eines Clusters
- Verwenden eines Hadoop-JAR-Schritts mit `command-runner.jar`
- Anmelden am Master-Knoten des Hadoop-Clusters und verwenden der Befehlszeilen-Clients `hadoop key` und `hdfs crypto`
- Verwenden der REST-APIs für Hadoop-KMS und HDFS

Weitere Informationen über die REST-APIs finden Sie in der jeweiligen Dokumentation für Hadoop-KMS und HDFS.

Erstellen von Verschlüsselungszonen und deren Schlüssel während der Cluster-Erstellung unter Verwendung der CLI

Mithilfe der `hdfs-encryption-zones`-Klassifizierung im Konfiguration-API-Vorgang können Sie bei der Erstellung eines Clusters einen Schlüsselnamen und eine Verschlüsselungszone angeben. Amazon EMR erstellt diesen Schlüssel im Hadoop-KMS auf Ihrem Cluster und konfiguriert die Verschlüsselungszone.

- Erstellen Sie einen Cluster mit dem folgenden Befehl.

```
aws emr create-cluster --release-label emr-7.1.0 --instance-type m5.xlarge --
instance-count 2 \
--applications Name=App1 Name=App2 --configurations https://s3.amazonaws.com/
mybucket/myfolder/myConfig.json
```

 Note

Linux-Zeilenumbruchzeichen (\) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (^).


myConfig.json:

```
[
  {
    "Classification": "hdfs-encryption-zones",
    "Properties": {
      "/myHDFSPath1": "path1_key",
      "/myHDFSPath2": "path2_key"
    }
  }
]
```

Manuelles Erstellen von Verschlüsselungszonen und deren Schlüssel auf dem Master-Knoten

1. Starten Ihres Clusters mit einer Amazon-EMR-Version größer als 4.1.0.
2. Stellen Sie mit dem Hauptknoten Ihres Clusters mithilfe von SSH eine Verbindung her.
3. Erstellen Sie einen Schlüssel im Hadoop-KMS.

```
$ hadoop key create path2_key
path2_key has been successfully created with options Options{cipher='AES/CTR/
NoPadding', bitLength=256, description='null', attributes=null}.
KMSClientProvider[http://ip-x-x-x-x.ec2.internal:16000/kms/v1/] has been updated.
```

 Important

Die Schlüsselnamen müssen im Hadoop-KMS in Kleinbuchstaben geschrieben werden. Wenn Sie einen Schlüssel mit Großbuchstaben verwenden, wird der Start Ihres Clusters fehlschlagen.

4. Erstellen des Verschlüsselungszonenpfads in HDFS.



```
$ hadoop fs -mkdir /myHDFSPath2
```

5. Machen Sie den HDFS-Pfad mit dem von Ihnen erstellten Schlüssel zu einer Verschlüsselungszone.

```
$ hdfs crypto -createZone -keyName path2_key -path /myHDFSPath2  
Added encryption zone /myHDFSPath2
```

Um Verschlüsselungszonen und ihre Schlüssel manuell zu erstellen, verwenden Sie AWS CLI

- Fügen Sie manuell Schritte zum Erstellen der KMS-Schlüssel und Verschlüsselungszonen mithilfe des folgenden Befehls hinzu.

```
aws emr add-steps --cluster-id j-2AXXXXXXGAPLF --steps Type=CUSTOM_JAR,Name="Create  
  First Hadoop KMS Key",Jar="command-runner.jar",ActionOnFailure=CONTINUE,Args=[/  
bin/bash,-c,"\"hadoop key create path1_key\""] \  
Type=CUSTOM_JAR,Name="Create First Hadoop HDFS Path",Jar="command-  
runner.jar",ActionOnFailure=CONTINUE,Args=[/bin/bash,-c,"\"hadoop fs -mkdir /  
myHDFSPath1\""] \  
Type=CUSTOM_JAR,Name="Create First Encryption Zone",Jar="command-  
runner.jar",ActionOnFailure=CONTINUE,Args=[/bin/bash,-c,"\"hdfs crypto -createZone  
  -keyName path1_key -path /myHDFSPath1\""] \  
Type=CUSTOM_JAR,Name="Create Second Hadoop KMS Key",Jar="command-  
runner.jar",ActionOnFailure=CONTINUE,Args=[/bin/bash,-c,"\"hadoop key create  
  path2_key\""] \  
Type=CUSTOM_JAR,Name="Create Second Hadoop HDFS Path",Jar="command-  
runner.jar",ActionOnFailure=CONTINUE,Args=[/bin/bash,-c,"\"hadoop fs -mkdir /  
myHDFSPath2\""] \  
Type=CUSTOM_JAR,Name="Create Second Encryption Zone",Jar="command-  
runner.jar",ActionOnFailure=CONTINUE,Args=[/bin/bash,-c,"\"hdfs crypto -createZone  
  -keyName path2_key -path /myHDFSPath2\""]
```

#### Note

Linux-Zeilenfortsetzungszeichen (\) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (^).

## Überlegungen zur transparenten HDFS-Verschlüsselung

Eine bewährte Methode ist das Erstellen einer Verschlüsselungszone für jede Anwendung, in der möglicherweise Dateien geschrieben werden. Sie können auch das gesamte HDFS verschlüsseln, indem Sie die `hdfs-encryption-zones` Klassifizierung in der Konfigurations-API verwenden und den Stammpfad (`/`) als Verschlüsselungszone angeben.

### Hadoop-Schlüsselverwaltungsserver (KMS)

[Hadoop](#) ist ein Schlüsselverwaltungsserver, der die Möglichkeit zur Implementierung kryptografischer Services für Hadoop-Cluster bietet und als wichtiger Anbieter von [Transparente Verschlüsselung in HDFS auf Amazon EMR](#) dient. Hadoop-KMS in Amazon EMR wird standardmäßig installiert und aktiviert, wenn Sie die Hadoop-Anwendung während des Starts eines EMR-Clusters auswählen. Hadoop KMS speichert die Schlüssel nicht selbst, mit Ausnahme des temporären Caching. Hadoop-KMS fungiert als Proxy zwischen dem Schlüsselanbieter und dem Client-Verwalter als Unterstützungsschlüsselspeicher – es ist jedoch kein Schlüsselspeicher. Der Standard-Keystore, der für Hadoop KMS erstellt wird, ist die Java Cryptography Extension (JCEKS). KeyStore Auch die JCE Unlimited Strength Policy ist inbegriffen, sodass Sie Schlüssel in der gewünschten Länge erstellen können. Hadoop-KMS unterstützt auch eine Reihe von ACLs, die den Zugriff auf Schlüssel und Schlüsselvorgänge unabhängig von anderen Client-Anwendungen wie beispielsweise HDFS kontrollieren. Die standardmäßige Schlüssellänge in Amazon EMR beträgt 256 Bit.

Verwenden Sie zur Konfiguration von Hadoop KMS die Klassifizierung, um Einstellungen zu ändern. `hadoop-kms-site` Verwenden Sie die `classification kms-acls`, um ACLs zu konfigurieren.

Weitere Informationen finden Sie in der [Hadoop-KMS-Dokumentation](#). Hadoop KMS wird für die transparente HDFS-Verschlüsselung verwendet. Weitere Informationen über die transparente HDFS-Verschlüsselung finden Sie im Abschnitt [Transparente HDFS-Verschlüsselung](#) in der Apache-Hadoop-Dokumentation.

#### Note

In Amazon EMR ist KMS über HTTPS nicht standardmäßig im Hadoop-KMS aktiviert. Informationen zum Aktivieren von KMS über HTTPS finden Sie in der [Hadoop-KMS-Dokumentation](#).

**⚠ Important**

Die Schlüsselnamen müssen im Hadoop-KMS in Kleinbuchstaben geschrieben werden. Wenn Sie einen Schlüssel mit Großbuchstaben verwenden, wird der Start Ihres Clusters fehlschlagen.

## Konfigurieren von Hadoop KMS in Amazon EMR

Mit Amazon-EMR-Version 4.6.0 oder höher lautet `kms-http-port` 9700 und `kms-admin-port` 9701.

Sie können die Hadoop-KMS während der Cluster-Erstellung konfigurieren. Verwenden Sie dazu die Konfigurations-API für Amazon-EMR-Versionen. Nachstehend finden Sie die Konfigurationsobjektklassifizierungen, die für Hadoop-KMS zur Verfügung stehen:

### Konfigurationsklassifizierungen für Hadoop-KMS

Klassifizierung	Dateiname
hadoop-kms-site	kms-site.xml
hadoop-kms-acls	kms-acls.xml
hadoop-kms-env	kms-env.sh
hadoop-kms-log4j	kms-log4j.properties

### Einrichten der Hadoop-KMS-ACLs mithilfe von CLI

- Verwenden Sie den folgenden Befehl, um einen Cluster mithilfe des Hadoop KMS mit ACLs zu erstellen:

```
aws emr create-cluster --release-label emr-7.1.0 --instance-type m5.xlarge --
instance-count 2 \
--applications Name=App1 Name=App2 --configurations https://s3.amazonaws.com/
mybucket/myfolder/myConfig.json
```

**Note**

Linux-Zeilenumbruchzeichen (\) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (^).

myConfig.json:

```
[
  {
    "Classification": "hadoop-kms-acls",
    "Properties": {
      "hadoop.kms.blacklist.CREATE": "hdfs,foo,myBannedUser",
      "hadoop.kms.acl.ROLLOVER": "myAllowedUser"
    }
  }
]
```

Deaktivieren des Hadoop-KMS-Cache mithilfe der CLI

- Verwenden Sie den folgenden Befehl, um einen Cluster mit Hadoop-KMS zu erstellen, wobei `hadoop.kms.cache.enable` auf `false` festgelegt ist:

```
aws emr create-cluster --release-label emr-7.1.0 --instance-type m5.xlarge --
instance-count 2 \
--applications Name=App1 Name=App2 --configurations https://s3.amazonaws.com/
mybucket/myfolder/myConfig.json
```

**Note**

Linux-Zeilenumbruchzeichen (\) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (^).

myConfig.json:

```
[
  {
    "Classification": "hadoop-kms-site",
    "Properties": {
      "hadoop.kms.cache.enable": "false"
    }
  }
]
```

## Einrichten der Umgebungsvariablen im Skript **kms-env.sh** mit CLI

- Ändern der Einstellungen in `kms-env.sh` über die `hadoop-kms-env`-Konfiguration. Verwenden Sie den folgenden Befehl, um einen Cluster mit Hadoop-KMS zu erstellen:

```
aws emr create-cluster --release-label emr-7.1.0 --instance-type m5.xlarge --
instance-count 2 \
--applications Name=App1 Name=App2 --configurations https://s3.amazonaws.com/
mybucket/myfolder/myConfig.json
```

### Note

Linux-Zeilenfortsetzungszeichen (`\`) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (`^`).

## myConfig.json:

```
[
  {
    "Classification": "hadoop-kms-env",
    "Properties": {
    },
    "Configurations": [
      {
        "Classification": "export",
        "Properties": {
          "JAVA_LIBRARY_PATH": "/path/to/files",
          "KMS_SSL_KEYSTORE_FILE": "/non/Default/Path/.keystore",

```

```

        "KMS_SSL_KEYSTORE_PASS": "myPass"
    },
    "Configurations": [
    ]
}
]
}
]

```

Weitere Informationen zur Konfiguration von Hadoop-KMS finden Sie in der [Hadoop-KMS-Dokumentation](#).

## Transparente HDFS-Verschlüsselung auf EMR-Clustern mit mehreren Hauptknoten

[Apache Ranger](#) KMS wird in einem Amazon-EMR-Cluster mit mehreren Primärknoten für transparente Verschlüsselung in HDFS verwendet.

Apache Ranger KMS speichert seinen Root-Schlüssel und seine Encryption Zone (EZ)-Schlüssel in Ihrem Amazon RDS für einen Amazon-EMR-Cluster mit mehreren Primärknoten. Um die transparente Verschlüsselung in HDFS auf einem Amazon-EMR-Cluster mit mehreren Primärknoten zu aktivieren, müssen Sie die folgenden Konfigurationen bereitstellen.

- Amazon RDS oder Ihre eigene MySQL-Server-Verbindungs-URL zum Speichern des Ranger-KMS-Root-Schlüssels und des EZ-Schlüssels
- Benutzername und Passwort für MySQL
- Passwort für den Ranger-KMS-Root-Schlüssel
- Certificate Authority (CA) PEM-Datei für SSL-Verbindung mit MySQL-Server

Sie können diese Konfigurationen mithilfe der `ranger-kms-dbks-site`-Klassifizierung und der `ranger-kms-db-ca`-Klassifizierung bereitstellen, wie im folgenden Beispiel veranschaulicht.

```

[
  {
    "Classification": "ranger-kms-dbks-site",
    "Properties": {
      "ranger.ks.jpa.jdbc.url": "jdbc:log4jdbc:mysql://mysql-host-url.xxx-xxx-1.xxx.amazonaws.com:3306/rangerkms",

```

```

    "ranger.ks.jpa.jdbc.user": "mysql-user-name",
    "ranger.ks.jpa.jdbc.password": "mysql-password",
    "ranger.db.encrypt.key.password": "password-for-encrypting-a-master-key"
  }
},
{
  "Classification": "ranger-kms-db-ca",
  "Properties": {
    "ranger.kms.trust.ca.file.s3.url": "s3://rds-downloads/rds-ca-2019-root.pem"
  }
}
]

```

Nachfolgend finden Sie Konfigurationsobjektklassifizierungen für Apache Ranger KMS.

### Konfigurationsklassifizierungen für Hadoop-KMS

Klassifizierung	Beschreibung
ranger-kms-dbks-site	Ändert die Werte in der dbks-site.xml-Datei von Ranger KMS.
ranger-kms-site	Ändern Sie die Werte in der ranger-kms-site XML-Datei von Ranger KMS.
ranger-kms-env	Ändert die Werte in der Ranger KMS-Umgebung.
ranger-kms-log4j	Ändert die Werte in der kms-log4j.properties-Datei von Ranger KMS.
ranger-kms-db-ca	Ändert die Werte für die CA-Datei auf S3 für die MySQL SSL-Verbindung mit Ranger KMS.

### Überlegungen

- Es wird dringend empfohlen, Ihre Amazon-RDS-Instance zu verschlüsseln, um die Sicherheit zu verbessern. Weitere Informationen finden Sie unter [Übersicht über die Verschlüsselung von Amazon-RDS-Ressourcen](#).

- Es wird dringend empfohlen, für jeden Amazon-EMR-Cluster eine separate MySQL-Datenbank mit mehreren Primärknoten zu verwenden, um eine hohe Sicherheit zu gewährleisten.
- Um transparente Verschlüsselung in HDFS auf einem Amazon-EMR-Cluster mit mehreren Primärknoten zu konfigurieren, müssen Sie beim Erstellen des Clusters die `hdfs-encryption-zones`-Klassifizierung angeben. Andernfalls wird Ranger KMS nicht konfiguriert oder gestartet. Die Neukonfiguration der `hdfs-encryption-zones`-Klassifizierung oder einer der Hadoop KMS-Konfigurationsklassifizierungen auf einem laufenden Cluster wird auf Amazon-EMR-Clustern mit mehreren Primärknoten nicht unterstützt.

## Erstellen oder Ausführen einer Hadoop-Anwendung

### Themen

- [Erstellen von Binärdateien mit Amazon EMR](#)
- [Daten per Streaming verarbeiten](#)
- [Daten mit einer benutzerdefinierten JAR-Datei verarbeiten](#)

## Erstellen von Binärdateien mit Amazon EMR

Sie können Amazon EMR als Build-Umgebung zum Kompilieren von Programmen für die Verwendung in Ihrem Cluster nutzen. Programme, die Sie mit Amazon EMR verwenden, müssen auf einem System kompiliert werden, das dieselbe Linux-Version wie Amazon EMR verwendet. Bei einer 32 Bit-Version sollten Sie die Kompilierung auf einem 32 Bit-Computer oder mit aktivierter 32 Bit-Cross-Compilation-Option durchgeführt haben. Bei einer 64 Bit-Version müssen Sie die Kompilierung auf einem 64 Bit-Computer oder mit aktivierter 64 Bit-Cross-Compilation-Option durchgeführt haben. Weitere Informationen zu EC2-Instance-Versionen finden Sie unter [Planen und Konfigurieren von EC2-Instances](#) im Verwaltungshandbuch für Amazon EMR. Unterstützte Programmiersprachen sind unter anderem C++, Python und C #.

In der folgenden Tabelle werden die Schritte für das Erstellen und Testen Ihrer Anwendung mit Amazon EMR aufgeführt.

### Prozess zum Erstellen eines Moduls

- 1 Stellen Sie eine Verbindung mit dem Master-Knoten Ihres Clusters her.
- 2 Kopieren Sie Quelldateien auf den Master-Knoten.



- 3 Erstellen Sie Binärdateien mit den erforderlichen Optimierungen.
- 4 Kopieren Sie die Binärdateien vom Hauptknoten zu Amazon S3.

Die Einzelheiten für jeden dieser Schritte finden Sie in den folgenden Abschnitten.

So verbinden Sie sich mit dem Master-Knoten des Clusters

- Befolgen Sie die Anweisungen unter [Verbindung mit dem Hauptknoten mit SSH](#) im Verwaltungshandbuch für Amazon EMR.

So kopieren Sie Quelldateien auf den Master-Knoten

1. Platzieren Sie Ihre Quelldateien in einem Amazon-S3-Bucket. Weitere Informationen zum Erstellen von Buckets und zum Verschieben von Daten nach Amazon S3 finden Sie im [Benutzerhandbuch für den Amazon Simple Storage Service](#).
2. Erstellen Sie einen Ordner für Ihre Quelldateien auf Ihrem Hadoop-Cluster, indem Sie einen Befehl wie den folgenden eingeben:

```
mkdir SourceFiles
```

3. Kopieren Sie Ihre Quelldateien von Amazon S3, auf den Hauptknoten, indem Sie einen Befehl wie den folgenden eingeben:

```
hadoop fs -get s3://mybucket/SourceFiles SourceFiles
```

Erstellen von Binärdateien mit den erforderlichen Optimierungen

Die Erstellung Ihrer Binärdateien hängt von vielen Faktoren ab. Befolgen Sie beim Einrichten und Konfigurieren Ihrer Umgebung die Anweisungen zu Ihren Build-Tools. Sie können Hadoop-System-Specification-Befehle nutzen, um Cluster-Informationen zur Definition der Installation Ihrer Build-Umgebung abzurufen.

So ermitteln Sie Systemspezifikationen

- Verwenden Sie die folgenden Befehle, um die zum Erstellen Ihrer Binärdateien verwendete Architektur abzurufen.

- a. Um die Debian-Version abzurufen, geben Sie den folgenden Befehl ein:

```
master$ cat /etc/issue
```

Die Ausgabe sieht folgendermaßen oder ähnlich aus.

```
Debian GNU/Linux 5.0
```

- b. Um den öffentlichen DNS-Namen und die Prozessorgöße abzurufen, geben Sie den folgenden Befehl ein:

```
master$ uname -a
```

Die Ausgabe sieht folgendermaßen oder ähnlich aus.

```
Linux domU-12-31-39-17-29-39.compute-1.internal 2.6.21.7-2.fc8xen #1 SMP Fri  
Feb 15 12:34:28 EST 2008 x86_64 GNU/Linux
```

- c. Um die Prozessorgeschwindigkeit abzurufen, geben Sie den folgenden Befehl ein:

```
master$ cat /proc/cpuinfo
```

Die Ausgabe sieht folgendermaßen oder ähnlich aus.

```
processor : 0  
vendor_id : GenuineIntel  
model name : Intel(R) Xeon(R) CPU E5430 @ 2.66GHz  
flags : fpu tsc msr pae mce cx8 apic mca cmov pat pse36 clflush dts acpi mmx  
fxsr sse sse2 ss ht tm syscall nx lm constant_tsc pni monitor ds_cpl vmx est  
tm2 ssse3 cx16 xtpr cda lahf_lm  
...
```

Sobald Ihre Binärdateien erstellt sind, können Sie die Dateien in Amazon S3 kopieren.

So kopieren Sie die Binärdateien vom Hauptknoten zu Amazon S3

- Geben Sie den folgenden Befehl ein, um die Binärdateien in Ihren Amazon-S3-Bucket zu kopieren:

```
hadoop fs -put BinaryFiles s3://mybucket/BinaryDestination
```

## Daten per Streaming verarbeiten

Hadoop-Streaming ist ein im Lieferumfang von Hadoop integriertes Hilfsprogramm, mit dem Sie MapReduce ausführbare Dateien in anderen Sprachen als Java entwickeln können. Streaming wird als JAR-Datei implementiert. Somit können Sie die Datei wie jede andere JAR-Datei über die Amazon-EMR-API oder die Befehlszeile ausführen.

In diesem Abschnitt wird die Verwendung von Streaming mit Amazon EMR beschrieben.

### Note

Apache Hadoop Streaming ist ein unabhängiges Tool. Daher werden seine Funktionen und Parameter hier nicht beschrieben. [Weitere Informationen zum Hadoop-Streaming finden Sie unter http://hadoop.apache.org/docs/stable/hadoop-streaming/.html](http://hadoop.apache.org/docs/stable/hadoop-streaming/.html). [HadoopStreaming](#)

## Verwenden des Hadoop-Streaming-Dienstprogramms

In diesem Abschnitt wird die Verwendung des Hadoop-Streaming-Dienstprogramms beschrieben.

### Hadoop-Prozess

- 1 Schreiben Sie Ihre ausführbare Mapper- und Reducer-Datei in der Programmiersprache Ihrer Wahl.  
  
Folgen Sie den Anweisungen in der Dokumentation zu Hadoop, um die ausführbaren Dateien zu schreiben. Die Programme sollten ihre Eingaben über die Standardeingabe lesen und die Daten über die Standardausgabedaten ausgeben. Standardmäßig stellt jede Zeile der Eingabe/Ausgabe einen Datensatz dar. Der erste Tabulator jeder Zeile dient als Trennzeichen zwischen Schlüssel und Wert.
- 2 Testen Sie Ihre ausführbaren Dateien lokal und laden Sie sie in Amazon S3 hoch.
- 3 Verwenden Sie die Amazon EMR oder die Amazon-EMR-Konsole, um Ihre Anwendung auszuführen.

Jedes Mapper-Skript startet als separater Prozess im Cluster. Jede ausführbare Reducer-Datei wandelt die Ausgabe der ausführbaren Mapper-Datei in die Datenausgabe des Auftragverlaufs um.

Die Parameter `input`, `output`, `mapper` und `reducer` sind für die meisten Streaming-Anwendungen erforderlich. Die folgende Tabelle beschreibt diese und andere (optionale) Parameter.

Parameter	Beschreibung	Erforderlich
<code>-input</code>	<p>Speicherort der Eingabedaten in Amazon S3.</p> <p>Typ: Zeichenfolge</p> <p>Standard: keiner</p> <p>Beschränkung: URI. Wenn kein Protokoll angegeben wurde, wird das Standarddateisystem des Clusters verwendet.</p>	Ja
<code>-output</code>	<p>Speicherort in Amazon S3, in dem Amazon EMR die verarbeiteten Daten hochlädt.</p> <p>Typ: Zeichenfolge</p> <p>Standard: keiner</p> <p>Beschränkung: URI</p> <p>Standard: Wenn kein Speicherort angegeben ist, werden die Daten von Amazon EMR in den von <code>input</code> angegebenen Speicherort hochgeladen.</p>	Ja
<code>-mapper</code>	<p>Name der ausführbaren Mapper-Datei.</p> <p>Typ: Zeichenfolge</p> <p>Standard: keiner</p>	Ja
<code>-reducer</code>	<p>Name der ausführbaren Reducer-Datei.</p> <p>Typ: Zeichenfolge</p>	Ja

Parameter	Beschreibung	Erforderlich
	Standard: keiner	
-cacheFile	<p>Ein Amazon-S3-Speicherort mit Dateien, die Hadoop in Ihr lokales Arbeitsverzeichnis kopieren soll (in erster Linie zur Verbesserung der Leistung).</p> <p>Typ: Zeichenfolge</p> <p>Standard: keiner</p> <p>Beschränkungen: [URI]#[symlink-Name zur Erstellung im Arbeitsverzeichnis]</p>	Nein
-cacheArchive	<p>JAR-Datei, die in das Arbeitsverzeichnis extrahiert werden soll</p> <p>Typ: Zeichenfolge</p> <p>Standard: keiner</p> <p>Beschränkungen: [URI]#[symlink-Verzeichnisname zur Erstellung im Arbeitsverzeichnis]</p>	Nein
-combiner	<p>Kombiniert Ergebnisse</p> <p>Typ: Zeichenfolge</p> <p>Standard: keiner</p> <p>Einschränkungen: Java-Klassenname</p>	Nein

Das folgende Codebeispiel ist eine ausführbare Mapper-Datei, die in Python geschrieben ist. Dieses Skript ist Teil der WordCount Beispielanwendung.

```
#!/usr/bin/python
import sys

def main(argv):
```

```
line = sys.stdin.readline()
try:
    while line:
        line = line.rstrip()
        words = line.split()
        for word in words:
            print "LongValueSum:" + word + "\t" + "1"
        line = sys.stdin.readline()
except "end of file":
    return None
if __name__ == "__main__":
    main(sys.argv)
```

## Einen Streaming-Schritt übermitteln

In diesem Abschnitt werden die Grundlagen der Übermittlung eines Streaming-Schritts an einen Cluster behandelt. Eine Streaming-Anwendung liest Eingaben aus der Standardeingabe und führt dann ein Skript oder eine ausführbare Datei (der Mapper) für den Eingang aus. Das Ergebnis aus den Eingaben wird lokal gespeichert, in der Regel auf einer Hadoop Distributed File System (HDFS)-Partition. Nachdem alle Eingaben vom Mapper verarbeitet wurden, verarbeitet ein zweites Skript oder eine ausführbare Datei (der Reducer) die Ergebnisse des Mappers. Die Ergebnisse aus dem Reducer werden an die Standardausgabe gesendet. Sie können eine Reihe von Streaming-Schritten verketteten, wobei die Ausgaben eines Schritts zur Eingabe eines weiteren Schritts werden.

Der Mapper und der Reducer können als Datei oder als Java-Klasse referenziert werden. Sie können Mapper und Reducer in einer der unterstützten Sprachen implementieren (zum Beispiel Ruby, Perl, Python, PHP oder Bash).

### Einen Streaming-Schritt mithilfe der Konsole übermitteln

In diesem Beispiel wird beschrieben, wie Sie mithilfe der Amazon-EMR-Konsole einen Streaming-Schritt an einen Cluster übermitteln, der ausgeführt wird.

#### So übermitteln Sie einen Streaming-Schritt

1. Öffnen Sie die Amazon EMR-Konsole unter <https://console.aws.amazon.com/emr>.
2. Wählen Sie unter Cluster List (Clusterliste) den Namen des Clusters aus.
3. Scrollen Sie zum Abschnitt Steps (Schritte) und erweitern Sie ihn. Wählen Sie anschließend Add step (Schritt hinzufügen) aus.
4. Gehen Sie im Dialogfeld Add Step folgendermaßen vor:

- Wählen Sie unter Step type (Schritttyp) die Option Streaming program (Streaming-Programm) aus.
  - Übernehmen Sie unter Name den Standardnamen (Streaming program) oder geben Sie einen neuen Namen ein.
  - Geben Sie für Mapper den Speicherort Ihrer Mapper-Klasse in Hadoop oder einen S3-Bucket ein, in dem sich die ausführbare Mapper-Datei (z. B. ein Python-Programm) befindet. Der Pfadwert muss das Format *BucketName/path/MappeExecutabl*e haben.
  - Geben Sie für Reducer den Speicherort Ihrer Reducer-Klasse in Hadoop oder einen S3-Bucket ein, in dem sich die ausführbare Reducer-Datei (z. B. ein Python-Programm) befindet. Der Pfadwert muss das Format *BucketName/Pfad/*haben *MappeExecutabl*e. Amazon EMR unterstützt das spezielle Schlüsselwort aggregate. Weitere Informationen finden Sie in der Aggregate-Bibliothek von Hadoop.
  - Geben Sie unter Input S3 location (S3-Eingabespeicherort) den Speicherort Ihrer Eingabedaten ein oder browsen zum Speicherort der Daten.
  - Unter S3-Ausgabespeicherort geben Sie den Namen Ihres Amazon-S3-Ausgabe-Buckets ein oder navigieren zu dem entsprechenden S3-Bucket.
  - Lassen Sie unter Arguments (Parameter) das Feld leer.
  - Übernehmen Sie unter Action on failure (Aktion bei Fehler) die Standardeinstellung Continue (Fortfahren).
5. Wählen Sie Hinzufügen aus. Der Schritt wird in der Konsole mit dem Status "Pending" angezeigt.
  6. Bei Ausführung des Schritts ändert sich dessen Status von Pending zu Running und Completed. Wählen Sie das Symbol Refresh über der Spalte "Actions" aus, um den Status zu aktualisieren.

## AWS CLI

Diese Beispiele zeigen, wie Sie mithilfe des AWS CLI einen Cluster erstellen und einen Streaming-Schritt einreichen können.

Um einen Cluster zu erstellen und einen Streaming-Schritt einzureichen, verwenden Sie den AWS CLI

- Um einen Cluster zu erstellen und einen Streaming-Schritt mit dem zu senden AWS CLI, geben Sie den folgenden Befehl ein und ersetzen Sie *MyKey* durch den Namen Ihres EC2-Schlüsselpaars. Beachten Sie, dass Ihr Argument für `--files` der Amazon-S3-Pfad zum

Speicherort Ihres Skripts und die Argumente für `-mapper` und `-reducer` die Namen der jeweiligen Skriptdateien sein sollten.

```
aws emr create-cluster --name "Test cluster" --release-label emr-7.1.0 --
applications Name=Hue Name=Hive Name=Pig --use-default-roles \
--ec2-attributes KeyName=myKey --instance-type m5.xlarge --instance-count 3 \
--steps Type=STREAMING,Name="Streaming Program",ActionOnFailure=CONTINUE,Args=[--
files,pathtoscripts,-mapper,mapperscript,-reducer,reducerscript,aggregate,-
input,pathtoinputdata,-output,pathtooutputbucket]
```

#### Note

Linux-Zeilenfortsetzungszeichen (`\`) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (`^`).

Wenn Sie die Instance-Anzahl ohne den Parameter `--instance-groups` angeben, wird ein einzelner Master-Knoten gestartet. Die verbleibenden Instances werden als Core-Knoten aufgerufen. Alle Knoten verwenden den im Befehl angegebenen Instance-Typ.

#### Note

Wenn Sie zuvor nicht die standardmäßige Amazon-EMR-Servicerolle und das EC2-Instance-Profil erstellt haben, geben Sie `aws emr create-default-roles` ein, um sie zu erstellen, bevor Sie den Unterbefehl `create-cluster` eingeben.

Weitere Informationen zur Verwendung von Amazon EMR-Befehlen finden Sie AWS CLI unter <https://docs.aws.amazon.com/cli/latest/reference/emr>.

## Daten mit einer benutzerdefinierten JAR-Datei verarbeiten

Eine benutzerdefinierte JAR-Datei führt ein kompiliertes Java-Programm aus, das Sie in Amazon S3 hochladen. Sie sollten das Programm mit der Hadoop-Version kompilieren, die Sie starten möchten, und senden Sie einen `CUSTOM_JAR`-Schritt an Ihren Amazon-EMR-Cluster. Weitere Informationen zum Kompilieren einer JAR-Datei finden Sie unter [Erstellen von Binärdateien mit Amazon EMR](#).



Weitere Informationen zum Erstellen einer MapReduce Hadoop-Anwendung finden Sie im [MapReduce Tutorial](#) in der Apache Hadoop-Dokumentation.

## Themen

- [Einen benutzerdefinierten JAR-Schritt übermitteln](#)

## Einen benutzerdefinierten JAR-Schritt übermitteln

Eine benutzerdefinierte JAR-Datei führt ein kompiliertes Java-Programm aus, das Sie in Amazon S3 hochladen. Sie sollten das Programm mit der Hadoop-Version kompilieren, die Sie starten möchten, und senden Sie einen CUSTOM\_JAR-Schritt an Ihren Amazon-EMR-Cluster. Weitere Informationen zum Kompilieren einer JAR-Datei finden Sie unter [Erstellen von Binärdateien mit Amazon EMR](#).

Weitere Informationen zum Erstellen einer MapReduce Hadoop-Anwendung finden Sie im [MapReduce Tutorial](#) in der Apache Hadoop-Dokumentation.

In diesem Abschnitt werden die Grundlagen der Übermittlung eines benutzerdefinierten JAR-Schritts in Amazon EMR behandelt. Wenn Sie einen benutzerdefinierten JAR-Schritt übermitteln, können Sie mit der Java-Programmiersprache ein Skript schreiben, das Ihre Daten verarbeitet.

## Einen benutzerdefinierten JAR-Schritt mithilfe der Konsole übermitteln

In diesem Beispiel wird beschrieben, wie Sie mithilfe der Amazon-EMR-Konsole einen benutzerdefinierten JAR-Schritt an einen aktiven Cluster senden.

So übermitteln Sie einen benutzerdefinierten JAR-Schritt mithilfe der Konsole

1. Öffnen Sie die Amazon EMR-Konsole unter <https://console.aws.amazon.com/emr>.
2. Wählen Sie unter Cluster List (Clusterliste) den Namen des Clusters aus.
3. Scrollen Sie zum Abschnitt Steps (Schritte) und erweitern Sie ihn. Wählen Sie anschließend Add step (Schritt hinzufügen) aus.
4. Gehen Sie im Dialogfeld Schritt hinzufügen folgendermaßen vor:
  - Wählen Sie Custom JAR (Benutzerdefiniertes JAR) als Step type (Schritttyp) aus.
  - Übernehmen Sie unter Name den Standardnamen (Custom JAR) oder geben Sie einen neuen Namen ein.

- Geben Sie unter JAR S3 location (JAR S3-Speicherort) den Speicherort Ihrer JAR-Datei ein oder navigieren Sie dorthin. Der JAR-Speicherort kann ein Pfad in S3 oder eine voll qualifizierte Java-Klasse im Klassenpfad sein.
  - Geben Sie unter Arguments (Argumente) Ihre Argumente als durch Leerzeichen getrennte Zeichenfolgen ein oder lassen Sie das Feld leer.
  - Übernehmen Sie unter Action on failure (Aktion bei Fehler) die Standardeinstellung Continue (Fortfahren).
5. Wählen Sie Hinzufügen aus. Der Schritt wird in der Konsole mit dem Status "Pending" angezeigt.
  6. Bei Ausführung des Schritts ändert sich dessen Status von Pending zu Running und Completed. Wählen Sie das Symbol Refresh über der Spalte "Actions" aus, um den Status zu aktualisieren.

### Starten eines Clusters und Einreichen eines benutzerdefinierten JAR-Schritts mit dem AWS CLI

Um einen Cluster zu starten und einen benutzerdefinierten JAR-Schritt einzureichen, verwenden Sie AWS CLI

Um einen Cluster zu starten und einen benutzerdefinierten JAR-Schritt mit dem zu senden AWS CLI, geben Sie den `create-cluster` Unterbefehl mit dem `--steps` Parameter ein.

- Um einen Cluster zu starten und einen benutzerdefinierten JAR-Schritt zu übermitteln, geben Sie den folgenden Befehl ein. Ersetzen Sie *myKey* durch den Namen Ihres EC2-Schlüsselpaares und *mybucket* durch Ihren Bucket-Namen.

```
aws emr create-cluster --name "Test cluster" --release-label emr-7.1.0 \
--applications Name=Hue Name=Hive Name=Pig --use-default-roles \
--ec2-attributes KeyName=myKey --instance-type m5.xlarge --instance-count 3 \
--steps Type=CUSTOM_JAR,Name="Custom JAR
Step",ActionOnFailure=CONTINUE,Jar=pathtojarfile,Args=["pathtoinputdata","pathtooutputbucket"]
```

#### Note

Linux-Zeilenfortsetzungszeichen (\) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (^).

Wenn Sie die Instance-Anzahl ohne den `--instance-groups`-Parameter angeben, wird ein einzelner Primärknoten gestartet. Die verbleibenden Instances werden dabei als Core-Knoten gestartet. Alle Knoten verwenden den im Befehl angegebenen Instance-Typ.

#### Note

Wenn Sie zuvor nicht die standardmäßige Amazon-EMR-Servicerolle und das EC2-Instance-Profil erstellt haben, geben Sie `aws emr create-default-roles` ein, um sie zu erstellen, bevor Sie den Unterbefehl `create-cluster` eingeben.

Weitere Informationen zur Verwendung von Amazon EMR-Befehlen finden Sie AWS CLI unter <https://docs.aws.amazon.com/cli/latest/reference/emr>.

## Abhängigkeiten von Drittanbietern

Manchmal kann es notwendig sein, JARs für die Verwendung mit Ihrem Programm in den MapReduce Klassenpfad aufzunehmen. Sie haben dafür zwei Möglichkeiten:

- Fügen Sie `--libjars s3://URI_to_JAR` in den Schrittoptionen für die Vorgehensweise in [Starten eines Clusters und Einreichen eines benutzerdefinierten JAR-Schritts mit dem AWS CLI](#) ein.
- Starten Sie den Cluster mit einer geänderten `mapreduce.application.classpath`-Einstellung in `mapred-site.xml`. Verwenden Sie die `mapred-site`-Konfigurationsklassifizierung. Um den Cluster mit dem Schritt `using` zu erstellen AWS CLI, würde das wie folgt aussehen:

```
aws emr create-cluster --release-label emr-7.1.0 \
--applications Name=Hue Name=Hive Name=Pig --use-default-roles \
--instance-type m5.xlarge --instance-count 2 --ec2-attributes KeyName=myKey \
--steps Type=CUSTOM_JAR,Name="Custom JAR  
Step",ActionOnFailure=CONTINUE,Jar=pathtojarfile,Args=["pathtoinputdata", "pathtooutputbucket  
\
--configurations https://s3.amazonaws.com/mybucket/myfolder/myConfig.json
```

myConfig.json:

```
[
```

```
{
  "Classification": "mapred-site",
  "Properties": {
    "mapreduce.application.classpath": "path1,path2"
  }
}
```

Die durch Kommas getrennte Liste der Pfade sollte jedem JVM-Klassenpfad für die Aufgabe angehängt werden.

## Aktivieren Sie die Erkennung ungleichmäßiger Speicherzugriffe für YARN-Container

Mit Amazon-EMR-Versionen 6.x und höher können Sie Non-Uniform Memory Access (NUMA) für die Mehrfachverarbeitung Ihrer Daten auf Clustern verwenden. NUMA ist ein Entwurfsmuster für Computerspeicher, bei dem der Prozessor schneller auf seinen eigenen lokalen Speicher zugreifen kann als auf Speicher auf einem anderen Prozessor oder auf den Speicher, der von mehreren Prozessoren gemeinsam genutzt wird. YARN-Container weisen eine bessere Leistung mit NUMA auf, da sie sich an einen bestimmten NUMA-Knoten binden können, der alle nachfolgenden Speicherzuweisungen bedient. Dadurch wird die Häufigkeit reduziert, mit der Ihr Cluster auf den Remotespeicher zugreifen muss.

Sie können die NUMA-Unterstützung für den YARN-Container aktivieren, wenn es sich bei der Worker-Knoten-Maschine um eine NUMA-Multi-Knoten-Maschine handelt. Führen Sie den folgenden Befehl aus, um zu überprüfen, ob es sich bei einem NUMA-Worker-Knoten um einen Single-NUMA- oder Multi-Knoten handelt.

```
lscpu | grep -i numa
NUMA node(s): 2
```

Im Allgemeinen haben Instances, die größer als 12x sind, zwei NUMA-Knoten. Dies gilt nicht für Metal-Instances..

So aktivieren Sie die Erkennung von NUMA-YARN-Containern

1. Verwenden Sie die folgende `yarn-site`-Konfiguration in Ihrem Amazon-EMR-6.x-Cluster.

```
[
  {
    "classification": "yarn-site",
    "properties": {
      "yarn.nodemanager.linux-container-executor.nonsecure-mode.local-
user": "yarn",
      "yarn.nodemanager.linux-container-executor.group": "yarn",
      "yarn.nodemanager.container-
executor.class": "org.apache.hadoop.yarn.server.nodemanager.LinuxContainerExecutor",
      "yarn.nodemanager.numa-awareness.enabled": "true",
      "yarn.nodemanager.numa-awareness.numactl.cmd": "/usr/bin/numactl",
      "yarn.nodemanager.numa-awareness.read-topology": "true"
    },
    "configurations": []
  }
]
```

2. Stellen Sie die folgende Bootstrap-Aktion in Ihrem Cluster bereit.

```
#!/bin/bash

sudo yum -y install numactl
echo 1 | sudo tee /proc/sys/kernel/numa_balancing

echo "banned.users=mapred,bin,hdfs" >> /etc/hadoop/conf/container-executor.cfg
rm -rf /var/log/hadoop-yarn/
sudo chown -R yarn:hadoop /var/log/hadoop-yarn/
sudo chmod 755 -R /var/log/hadoop-yarn/

sudo chmod 6050 /etc/hadoop/conf/container-executor.cfg

mkdir /mnt/yarn && sudo chmod 755 -R /mnt/yarn && sudo chown -R yarn:hadoop /mnt/
yarn
mkdir /mnt1/yarn && sudo chmod 755 -R /mnt1/yarn && sudo chown -R yarn:hadoop /
mnt1/yarn
mkdir /mnt2/yarn && sudo chmod 755 -R /mnt2/yarn && sudo chown -R yarn:hadoop /
mnt2/yarn
```

3. Jeder Container muss NUMA kennen. Sie können die virtuelle Java-Maschine (JVM) in jedem Container mit einem NUMA-Flag benachrichtigen. Um beispielsweise die JVM zu benachrichtigen, NUMA in einem MapReduce-Auftrag zu verwenden, fügen Sie die folgenden Eigenschaften in `mapred-site.xml` hinzu.

```
<property>
  <name>mapreduce.reduce.java.opts</name>
  <value>-XX:+UseNUMA</value>
</property>
<property>
  <name>mapreduce.map.java.opts</name>
  <value>-XX:+UseNUMA</value>
</property>
```

- Um zu überprüfen, ob NUMA aktiviert sind, durchsuchen Sie eine der NodeManager-Protokolldateien mit dem folgenden Befehl.

```
grep "NUMA resources allocation is enabled," *
```

Um zu überprüfen, NodeManager ob einem Container NUMA Knotenressourcen zugewiesen wurden, durchsuchen NodeManager Sie das Protokoll mit dem folgenden Befehl und ersetzen *<container\_id>* Sie es durch Ihre eigene Container-ID.

```
grep "NUMA node" | grep <container_id>
```

## Hadoop-Versionsverlauf

In der folgenden Tabelle sind die Version von Hadoop aufgeführt, die in jeder Release-Version von Amazon EMR enthalten ist, zusammen mit den Komponenten, die mit der Anwendung installiert wurden. Informationen zu den Komponentenversionen in den einzelnen Versionen finden Sie im Abschnitt [Komponentenversion für Ihre Version in Amazon-EMR-7.x-Versionen](#), [Amazon-EMR-6.x-Versionen](#) oder [Amazon-EMR-5.x-Versionen](#).

Die Hadoop-Versionsinformationen

Amazon-EMR-Version sbezeichnung	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.36.2	2.10.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-nameno

Amazon-EMR-Version sbezeichnung	Hadoop-Version	Mit Hadoop installierte Komponenten
		de, hadoop-https-server, hadoop-kms-server, hadoop- mapred, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server
emr-7.1.0	3.3.6	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop- hdfs-datanode, hadoop-hdfs- library, hadoop-hdfs-nameno de, hadoop-https-server, hadoop-kms-server, hadoop- mapred, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server
emr-7.0.0	3.3.6	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop- hdfs-datanode, hadoop-hdfs- library, hadoop-hdfs-nameno de, hadoop-https-server, hadoop-kms-server, hadoop- mapred, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server

Amazon-EMR-Version sbezeichnung	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-6.15.0	3.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.14.0	3.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server



Amazon-EMR-Version sbezeichnung	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-6.13.0	3.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.12.0	3.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon-EMR-Version sbezeichnung	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-6.11.1	3.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.11.0	3.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon-EMR-Version sbezeichnung	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-6.10.1	3.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.10.0	3.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon-EMR-Version sbezeichnung	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-6.9.1	3.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.9.0	3.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon-EMR-Version sbezeichnung	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-6.8.1	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.8.0	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon-EMR-Version sbezeichnung	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-6.7.0	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.36.1	2.10.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon-EMR-Version sbezeichnung	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.36.0	2.10.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.6.0	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon-EMR-Version sbezeichnung	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.35.0	2.10.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.5.0	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server



Amazon-EMR-Version sbezeichnung	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-6.4.0	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.3.1	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon-EMR-Version sbezeichnung	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-6.3.0	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.2.1	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon-EMR-Version sbezeichnung	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-6.2.0	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.1.1	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon-EMR-Version sbezeichnung	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-6.1.0	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.0.1	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon-EMR-Version sbezeichnung	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-6.0.0	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.34.0	2.10.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon-EMR-Version sbezeichnung	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.33.1	2.10.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.33.0	2.10.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon-EMR-Version sbezeichnung	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.32.1	2.10.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.32.0	2.10.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon-EMR-Version sbezeichnung	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.31.1	2.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.31.0	2.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server



Amazon-EMR-Version sbezeichnung	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.30.2	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.30.1	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon-EMR-Version sbezeichnung	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.30.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.29.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon-EMR-Version sbezeichnung	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.28.1	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.28.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon-EMR-Version sbezeichnung	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.27.1	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.27.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon-EMR-Version sbezeichnung	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.26.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.25.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon-EMR-Version sbezeichnung	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.24.1	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.24.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon-EMR-Version sbezeichnung	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.23.1	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.23.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon-EMR-Version sbezeichnung	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.22.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.21.2	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server



Amazon-EMR-Version sbezeichnung	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.21.1	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.21.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon-EMR-Version sbezeichnung	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.20.1	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.20.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon-EMR-Version sbezeichnung	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.19.1	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.19.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon-EMR-Version sbezeichnung	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.18.1	2.8.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.18.0	2.8.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon-EMR-Version sbezeichnung	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.17.2	2.8.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.17.1	2.8.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon-EMR-Version sbezeichnung	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.17.0	2.8.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.16.1	2.8.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon-EMR-Version sbezeichnung	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.16.0	2.8.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.15.1	2.8.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon-EMR-Version sbezeichnung	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.15.0	2.8.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.14.2	2.8.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server



Amazon-EMR-Version sbezeichnung	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.14.1	2.8.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.14.0	2.8.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon-EMR-Version sbezeichnung	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.13.1	2.8.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.13.0	2.8.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon-EMR-Version sbezeichnung	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.12.3	2.8.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.12.2	2.8.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon-EMR-Version sbezeichnung	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.12.1	2.8.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.12.0	2.8.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon-EMR-Version sbezeichnung	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.11.4	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.11.3	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon-EMR-Version sbezeichnung	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.11.2	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.11.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon-EMR-Version sbezeichnung	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.11.0	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.10.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon-EMR-Version sbezeichnung	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.10.0	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.9.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server



Amazon-EMR-Version sbezeichnung	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.9.0	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.8.3	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon-EMR-Version sbezeichnung	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.8.2	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.8.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon-EMR-Version sbezeichnung	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.8.0	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.7.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon-EMR-Version sbezeichnung	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.7.0	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.6.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon-EMR-Version sbezeichnung	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.6.0	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.5.4	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.5.3	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Amazon-EMR-Version sbezeichnung	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.5.2	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.5.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.5.0	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Amazon-EMR-Version sbezeichnung	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.4.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.4.0	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.3.2	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Amazon-EMR-Version sbezeichnung	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.3.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.3.0	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.2.3	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager



Amazon-EMR-Version sbezeichnung	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.2.2	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.2.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.2.0	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Amazon-EMR-Version sbezeichnung	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.1.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.1.0	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.0.3	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Amazon-EMR-Version sbezeichnung	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-5.0.2	2.7.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.0.1	2.7.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.0.0	2.7.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Amazon-EMR-Version sbezeichnung	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-4.9.6	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.9.5	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.9.4	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Amazon-EMR-Version sbezeichnung	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-4.9.3	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.9.2	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.9.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Amazon-EMR-Version sbezeichnung	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-4.8.5	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.8.4	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.8.3	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Amazon-EMR-Version sbezeichnung	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-4.8.2	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.8.1	2.7.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.8.0	2.7.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Amazon-EMR-Version sbezeichnung	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-4.7.4	2.7.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.7.3	2.7.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.7.2	2.7.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager



Amazon-EMR-Version sbezeichnung	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-4.7.1	2.7.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.7.0	2.7.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.6.1	2.7.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Amazon-EMR-Version sbezeichnung	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-4.6.0	2.7.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.5.0	2.7.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.4.0	2.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Amazon-EMR-Version sbezeichnung	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-4.3.0	2.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.2.0	2.6.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.1.0	2.6.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Amazon-EMR-Version sbezeichnung	Hadoop-Version	Mit Hadoop installierte Komponenten
emr-4.0.0	2.6.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datano-de, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

## Hadoop-Versionshinweise nach Version

### [Amazon EMR 6.6.0 – Hadoop-Versionshinweise](#)

#### Amazon EMR 6.6.0 – Hadoop-Versionshinweise

#### Amazon EMR 6.6.0 – Hadoop-Änderungen

Typ	Beschreibung
Fehler	Doppelte Datensätze beim Lesen von bZip2-Textdateien behoben.
Backport	<a href="#">HADOOP-18136</a> : Überprüfen Sie den <code>FileUtils.unTar()</code> -Umgang mit fehlenden <code>.tar</code> -Dateien
Backport	<a href="#">HADOOP-17627</a> : Backport auf Branch-3.2 HADOOP-17371, HADOOP-17621, HADOOP-17625, um Jetty auf 9.4.39 zu aktualisieren
Backport	<a href="#">HADOOP-17655</a> : Aktualisieren Sie Jetty auf 9.4.40

Typ	Beschreibung
Backport	<a href="#">HADOOP-17796</a> : Aktualisieren Sie Jetty auf 9.4.43
Backport	<a href="#">HADOOP-17661</a> : mvn versions:set kann pom.xml nicht analysieren
Backport	<a href="#">HADOOP-17236</a> : Erhöhen Sie Snakeyaml auf 1.26, um CVE-2017-18640 abzuschwächen
Backport	<a href="#">HADOOP-16717</a> : Entfernen Sie die IsLog4JLogger-Abhängigkeit von Log4 GenericsUtil jLoggerAdapter
Backport	<a href="#">HADOOP-17633</a> : Erhöhen Sie JSON-Smart aufgrund von CVEs auf 2.4.2 und auf 9.8 nimbus-jose-jwt
Backport	<a href="#">HADOOP-17844</a> : Aktualisieren Sie JSON Smart auf 2.4.7
Backport	<a href="#">HADOOP-17972</a> : <a href="#">Backport HADOOP-17683</a> (Update commons-io auf 2.8.0) für Branch-3.2
Backport	<a href="#">HADOOP-16555</a> : Commons-Compress auf Version 1.19 aktualisieren
Backport	<a href="#">HADOOP-17370</a> : Aktualisieren Sie commons-compress auf 1.21
Backport	<a href="#">HADOOP-17096</a> : StandardCompressor Korrigiert den Offset des Z-Eingabepuffers
Backport	<a href="#">HADOOP-17112</a> : Leerzeichen sind in Pfaden nicht erlaubt, wenn Dateien per Committer auf S3A gespeichert werden

Typ	Beschreibung
Backport	<a href="#">HADOOP-13500</a> : Synchronisieren der Iteration des Objekts mit den Konfigurationseigenschaften
Backport	<a href="#">HDFS-14099</a> : Unbekannter Frame-Deskriptor beim Dekomprimieren mehrerer Frames in ZStandardDecompressor
Backport	<a href="#">HDFS-16410</a> : Unsicheres XML-Parsen in OfflineEditsXmlLoader
Backport	<a href="#">HDFS-14498</a> : LeaseManager kann die Datei, deren Erstellung fehlgeschlagen ist, für immer wiederholen
Backport	<a href="#">HDFS-15290</a> : NPE ist beim Start aktiviert HttpServer NameNode
Backport	<a href="#">HDFS-15293</a> : Lockern Sie die Bedingungen für die Annahme von fsimage beim Empfang eines Checkpoints
Backport	<a href="#">HDFS-12979</a> : sollte nach dem Checkpointing hochgeladen werden StandbyNode FsImage ObserverNode
Backport	<a href="#">YARN-10538</a> : Fügen Sie der Liste der aktualisierten Knoten, die an den AM zurückgegeben wurden, Knoten für die Wiederinbetriebnahme hinzu
Backport	<a href="#">YARN-10472</a> : Backport von YARN-10314 (löst nur bei schattierten Client-Jars) nach Branch-3.2 YarnClient NoClassDefFoundError WebSocketException

Typ	Beschreibung
Backport	<a href="#">YARN-9968</a> : Public Localizer wird aufgrund von <code>NodeManager</code> in <code>NullPointerException</code> beendet
Backport	<a href="#">YARN-10651</a> : <code>CapacityScheduler</code> mit NPE in <code>AbstractYarnScheduler.updateNodeResource()</code> abgestürzt
Backport	<a href="#">YARN-9339</a> : Die Metrik für ausstehende Anwendungen ist falsch, nachdem die Anwendung in eine neue Warteschlange verschoben wurde
Backport	<a href="#">YARN-10438</a> : Behandeln von Null container Id in <code>ClientRMService#getContainerReport()</code>
Backport	<a href="#">YARN-7266</a> : ATS 1.5 kann nicht gestartet werden, wenn <code>RollingLevelDb</code> -Dateien beschädigt sind oder fehlen
Backport	<a href="#">YARN-9063</a> : ATS 1.5 kann nicht gestartet werden, wenn <code>RollingLevelDb</code> -Dateien beschädigt sind oder fehlen
Backport	<a href="#">YARN-9848</a> : Setzen Sie YARN-4946 zurück (RM sollte eine Anwendung nicht als <code>COMPLETED</code> betrachten, wenn sich die Protokollaggregation nicht in einem Endstatus befindet).

# Apache HBase

[HBase](#) ist eine nicht relationale, verteilte Open-Source-Datenbank, die im Rahmen des Hadoop-Projekts der Apache Software Foundation entwickelt wurde. HBase wird auf Hadoop Distributed File System (HDFS) ausgeführt, um für das Hadoop-Ökosystem nicht relationale Datenbankfunktionen bereitzustellen. HBase ist in Amazon-EMR-Version 4.6.0 und höher enthalten.

HBase arbeitet nahtlos mit Hadoop zusammen, teilt sein Dateisystem und dient als direkte Eingabe und Ausgabe für das MapReduce Framework und die Ausführungs-Engine. HBase lässt sich auch mit Apache Hive integrieren, ermöglicht SQL-artige Abfragen von HBase-Tabellen und Join-Vorgänge mit Hive-basierten Tabellen und unterstützt Java Database Connectivity (JDBC). Weitere Informationen zu HBase finden Sie unter [Apache HBase](#) und in der [HBase-Dokumentation](#) auf der Apache-Website. Ein Beispiel für die Verwendung von HBase mit Hive finden Sie im AWS Big Data-Blogbeitrag [Kombinieren von NoSQL und massively parallel analytics using Apache HBase and Apache Hive](#) on Amazon EMR.

Mit HBase auf Amazon EMR können Sie Ihre HBase-Daten auch direkt in Amazon Simple Storage Service Amazon S3 sichern und aus einer zuvor erstellten Sicherung wiederherstellen, wenn Sie einen HBase-Cluster starten. Amazon EMR bietet zusätzliche Optionen für die Integration Amazon S3 für eine dauerhafte Datenspeicherung und die Notfallwiederherstellung.

- HBase in Amazon S3 – Ab Amazon-EMR-Version 5.2.0 können Sie HBase auf Amazon S3 verwenden, um das HBase-Stammverzeichnis und die Metadaten eines Clusters direkt in Amazon S3 zu speichern. Anschließend können Sie einen neuen Cluster starten; dieser muss auf den Speicherort im Stammverzeichnis in Amazon S3 verweisen. Es kann nur jeweils ein Cluster den HBase-Standort in Amazon S3 verwenden, mit Ausnahme eines Lesereplikat-Clusters. Weitere Informationen finden Sie unter [HBase auf Amazon S3 \(Amazon-S3-Speichermodus\)](#).
- HBase-Lesereplikate – Amazon EMR ab Version 5.7.0 mit HBase auf Amazon S3 unterstützt Lesereplikat-Cluster. Ein Read Replica-Cluster bietet schreibgeschützten Zugriff auf die Dateien und Metadaten des Speichers eines primären Clusters, um schreibgeschützte Operationen durchzuführen. Weitere Informationen finden Sie unter [Verwenden eines Lesereplikat-Clusters](#).
- HBase-Snapshots - Als Alternative zu HBase auf Amazon S3 können Sie mit EMR ab Version 4.0 Snapshots Ihrer HBase-Daten direkt in Amazon S3 erstellen und dann Daten mithilfe der Snapshots wiederherstellen. Weitere Informationen finden Sie unter [Verwenden von HBase-Snapshots](#).



**⚠ Important**

Für die Amazon-EMR-HBase-Clusterskalierung empfehlen wir nicht, [verwaltete Skalierung](#) oder [Skalierung mit benutzerdefinierten Richtlinien](#) mit HBase-Clustern zu verwenden.

Die folgende Tabelle listet die Version von HBase auf, die in der neuesten Version der Amazon-EMR-7.x-Serie enthalten ist, zusammen mit den Komponenten, die Amazon EMR mit HBase installiert.

[Informationen zur Version der Komponenten, die in dieser Version mit HBase installiert wurden, finden Sie unter Komponentenversionen von Version 7.1.0.](#)

## HBase-Versionsinformationen für emr-7.1.0

Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-7.1.0	HBase 2.4.17	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-wal-cli, hadoop-client, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server

Die folgende Tabelle listet die Version von Tez auf, die in der neuesten Version der Amazon-EMR-6.x-Serie enthalten ist, zusammen mit den Komponenten, die Amazon EMR mit HBase installiert.

Die Version der Komponenten, die mit HBase in dieser Version installiert wurden, finden Sie unter [Komponentenversionen der Version 6.15.0](#).

#### HBase-Versionsinformationen für emr-6.15.0

Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-6.15.0	HBase 2.4.17	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-wal-cli, hadoop-client, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server

#### Note

Apache HBase HBCK2 ist ein separates Betriebssystem zur Reparatur von HBase-Regionen und Systemtabellen. In Amazon EMR Version 6.1.0 und höher wird die Datei `hbase-hbck2.jar` `/usr/lib/hbase-operator-tools/` auf dem primären Knoten bereitgestellt. Weitere Informationen zum Erstellen und Verwenden des Tools finden Sie unter [HBase HBCK2](#).

Die folgende Tabelle listet die Version von HBase auf, die in der neuesten Version der Amazon-EMR-5.x-Serie enthalten ist, zusammen mit den Komponenten, die Amazon EMR mit HBase installiert.

Informationen zur Version der Komponenten, die in dieser Version mit HBase installiert wurden, finden Sie unter [Komponentenversionen von Version 5.36.2](#).

HBase-Versionsinformationen für emr-5.36.2

Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-5.36.2	HBase 1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

## Themen

- [Erstellen eines Clusters mit HBase](#)
- [HBase auf Amazon S3 \(Amazon-S3-Speichermodus\)](#)
- [Write-Ahead-Protokolle \(WAL\) für Amazon EMR](#)
- [Verwenden der HBase-Shell](#)
- [Zugreifen auf HBase-Tabellen mit Hive](#)
- [Verwenden von HBase-Snapshots](#)
- [Konfigurieren von HBase](#)

- [Anzeigen der HBase-Benutzeroberfläche](#)
- [Anzeigen der HBase-Protokolldateien](#)
- [Überwachen von HBase mit Ganglia](#)
- [Migrieren von vorherigen HBase-Versionen](#)
- [Versionsverlauf von HBase](#)

## Erstellen eines Clusters mit HBase

Die Verfahren in diesem Abschnitt behandeln die Grundlagen des Starts eines Clusters mithilfe von und. AWS Management Console AWS CLI Ausführliche Informationen zur Planung, Konfiguration und Inbetriebnahme von Amazon EMR-Clustern finden Sie unter [Planen und Konfigurieren von Clustern](#) im Amazon EMR Management Guide.

### Erstellen eines Clusters mit HBase über die Konsole

Schnelle Schritte zum Starten von Clustern mit der Konsole finden Sie unter [Erste Schritte mit Amazon EMR](#) im Verwaltungshandbuch für Amazon EMR.

So starten Sie einen Cluster mit HBase über die Konsole

1. Öffnen Sie die Amazon EMR-Konsole unter <https://console.aws.amazon.com/emr>.
2. Wählen Sie Create Cluster (Cluster erstellen) und Go to advanced options (Zu erweiterten Optionen) aus.
3. Wählen Sie unter Software Configuration (Softwarekonfiguration) die Amazon Release Version 4.6.0 oder höher aus (wir empfehlen die neueste Version). Wählen Sie HBase und je nach Wunsch andere Anwendungen aus.
4. Wählen Sie ab Amazon-EMR-Version 5.2.0 unter HBase-Speichereinstellungen die Option HDFS oder S3. Weitere Informationen finden Sie unter [HBase auf Amazon S3 \(Amazon-S3-Speichermodus\)](#).
5. Wählen Sie nach Bedarf weitere Optionen und anschließend Create cluster (Cluster erstellen) aus.

### Erstellen eines Clusters mit HBase unter Verwendung des AWS CLI

Verwenden Sie den folgenden Befehl, um einen Cluster zu erstellen, auf dem HBase installiert ist:

```
aws emr create-cluster --name "Test cluster" --release-label emr-7.1.0 \  
--applications Name=HBase --use-default-roles --ec2-attributes KeyName=myKey \  
--instance-type m5.xlarge --instance-count 3
```

### Note

Linux-Zeilenfortsetzungszeichen (\) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (^).

Wenn Sie HBase auf Amazon S3 verwenden, geben Sie die Option `--configurations` mit einem Verweis auf ein JSON-Konfigurationsobjekt an. Das Konfigurationsobjekt muss eine `hbase-site`-Klassifizierung enthalten, die über die Eigenschaft `hbase.rootdir` den Speicherort in Amazon S3 angibt, in dem HBase-Daten gespeichert sind. Es muss außerdem eine `hbase`-Klassifizierung enthalten, die `s3` über die Eigenschaft `hbase.emr.storageMode` angibt. Das folgende Beispiel zeigt einen JSON-Codeausschnitt mit diesen Konfigurationseinstellungen.

```
[  
  {  
    "Classification": "hbase-site",  
    "Properties": {  
      "hbase.rootdir": "s3://MyBucket/MyHBaseStore"  
    }  
  },  
  {  
    "Classification": "hbase",  
    "Properties": {  
      "hbase.emr.storageMode": "s3"  
    }  
  }  
]
```

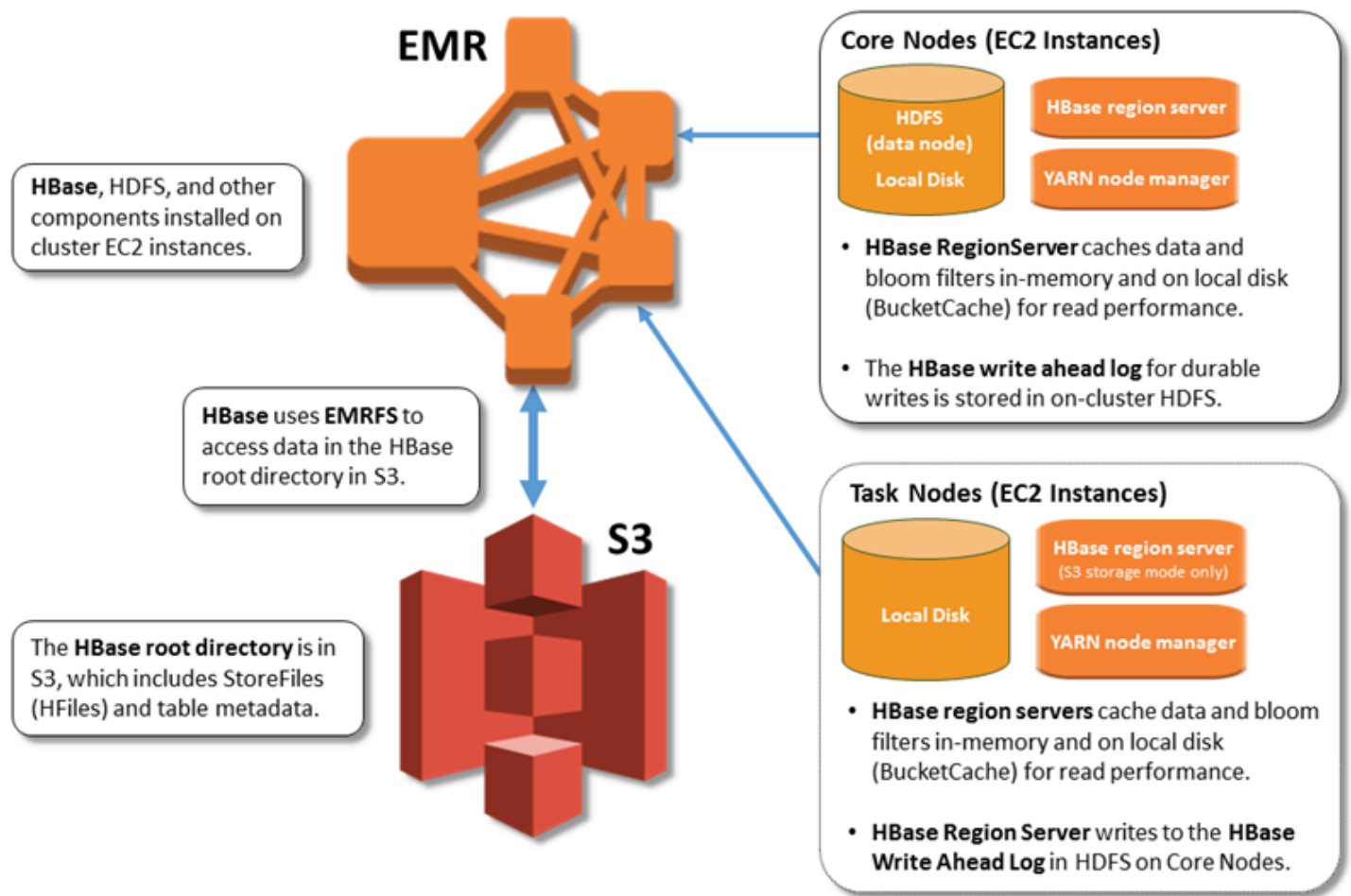
Weitere Informationen zu HBase auf Amazon S3 finden Sie unter [HBase auf Amazon S3 \(Amazon-S3-Speichermodus\)](#). Weitere Informationen zu Klassifizierungen finden Sie unter [Anwendungen konfigurieren](#).

## HBase auf Amazon S3 (Amazon-S3-Speichermodus)

Wenn Sie HBase auf Amazon-EMR-Version 5.2.0 oder höher ausführen, können Sie HBase auf Amazon S3 aktivieren. Dies bietet Ihnen folgende Vorteile:

- Das HBase-Stammverzeichnis wird in Amazon S3 gespeichert, einschließlich der HBase-Speicherdateien und Tabellenmetadaten. Diese Daten werden außerhalb des Clusters beibehalten und stehen für die Amazon-EC2-Availability-Zones zur Verfügung. Sie müssen keine Wiederherstellung unter Anwendung von Snapshots oder anderer Methoden durchführen.
- Mit Speicherdateien in Amazon S3 können Sie mit dreifacher Replikation in HDFS die Größe Ihres Amazon-EMR-Clusters an Ihre EDV-Anforderungen anpassen, nicht an die Datenanforderungen.
- Wenn Sie Amazon EMR ab Version 5.7.0 verwenden, können Sie einen Lesereplikat-Cluster einrichten. So können Sie schreibgeschützte Kopien der Daten in Amazon S3 pflegen. Sie können auf die Daten vom Read Replica-Cluster zugreifen, um gleichzeitig Lesevorgänge auszuführen und wenn der primäre Cluster nicht mehr verfügbar ist.
- In Amazon-EMR-Version 6.2.0 und höher verwendet persistentes HFile Tracking eine HBase-Systemtabelle namens `hbase:storefile`, die aufgerufen wird, um die für Lesevorgänge verwendeten HFile-Pfade direkt zu verfolgen. Dieses Feature ist standardmäßig aktiviert und erfordert keine manuelle Migration.

Die folgende Abbildung zeigt die HBase-Komponenten, die für HBase auf Amazon S3 relevant sind.



## Aktivieren von HBase auf Amazon S3

Sie können HBase auf Amazon S3 mithilfe der Amazon EMR-Konsole, der AWS CLI, oder der Amazon EMR-API aktivieren. Die Konfiguration wird während der Cluster-Erstellung als Option angeboten. Wenn Sie die Konsole verwenden, wählen Sie die Einstellung über **Advanced options** (Erweiterte Optionen) aus. In der AWS CLI verwenden Sie die Option `--configurations`, um ein JSON-Konfigurationsobjekt bereitzustellen. Die Eigenschaften des Konfigurationsobjekts geben den Speichermodus und das Stammverzeichnis in Amazon S3 an. Der Amazon-S3-Speicherort, den Sie angeben, sollte sich in derselben Region befinden wie Ihr Amazon-EMR-Cluster. Es kann nur jeweils ein aktiver Cluster das gleiche HBase-Stammverzeichnis in Amazon S3 verwenden. Anweisungen zur Konsole und ein detailliertes Beispiel zum Erstellen eines Clusters mithilfe von finden Sie unter [AWS CLI Erstellen eines Clusters mit HBase](#). Der folgende JSON-Codeausschnitt zeigt ein Beispiel für ein Konfigurationsobjekt.

```
{
```

```
"Classification": "hbase-site",
"Properties": {
  "hbase.rootdir": "s3://my-bucket/my-hbase-rootdir"}
},
{
  "Classification": "hbase",
  "Properties": {
    "hbase.emr.storageMode": "s3"
  }
}
```

### Note

Wenn Sie einen Amazon-S3-Bucket als `rootdir` für HBase verwenden, müssen Sie am Ende der Amazon-S3-URI einen Schrägstrich hinzufügen. Zum Beispiel müssen Sie statt `"hbase.rootdir: s3://my-bucket/"` `"hbase.rootdir: s3://my-bucket"` verwenden, um Probleme zu vermeiden.

## Verwenden eines Lesereplikat-Clusters

Nachdem Sie einen primären Cluster mit HBase auf Amazon S3 eingerichtet haben, können Sie einen Lesereplikat-Cluster erstellen und konfigurieren, der auf die gleichen Daten wie der primäre Cluster schreibgeschützten Zugriff gewährt. Dies ist hilfreich, wenn Sie gleichzeitigen Zugriff auf Abfragedaten brauchen oder einen unterbrechungsfreien Zugriff benötigen, wenn der primäre Cluster nicht mehr verfügbar ist. Das Lesereplikat-Feature ist ab Version 5.7.0 von Amazon EMR verfügbar.

Der primäre Cluster sowie der Read Replica-Cluster werden auf die gleiche Weise eingerichtet, mit einem wichtigen Unterschied. Beide verweisen auf denselben `hbase.rootdir`-Speicherort. Die `hbase`-Klassifizierung für den Read Replica-Cluster enthält jedoch die Eigenschaft `"hbase.emr.readreplica.enabled": "true"`.

Mit der oben genannten JSON-Klassifizierung für den primären Cluster sieht die Konfiguration für einen Read Replica-Cluster wie folgt aus:

```
{
  "Classification": "hbase-site",
  "Properties": {
```

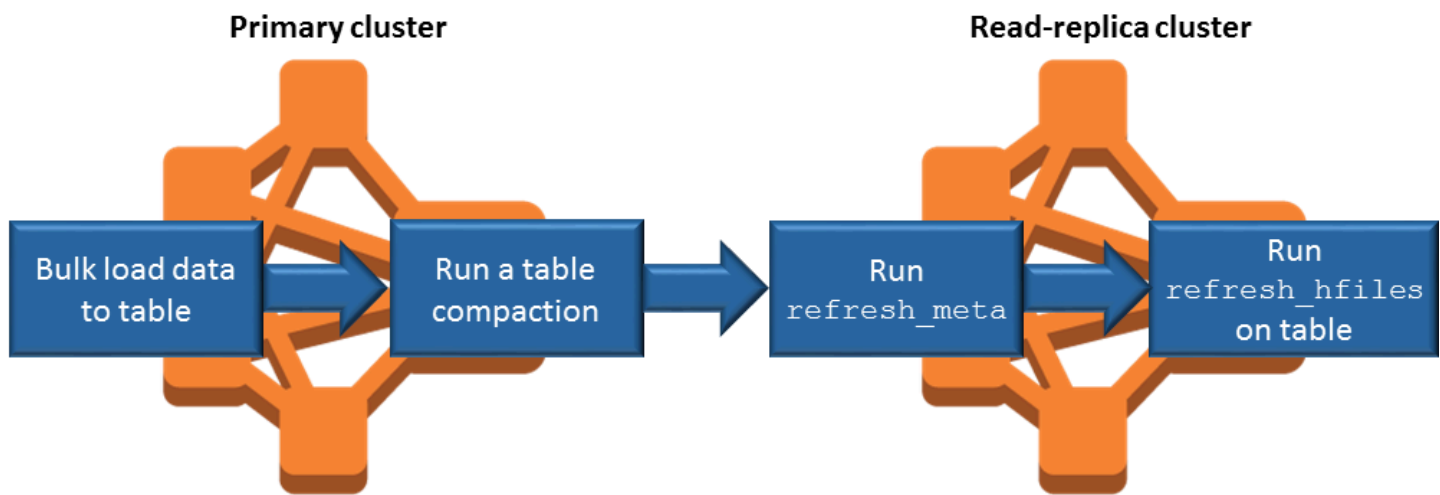


```
    "hbase.rootdir": "s3://my-bucket/my-hbase-rootdir"}
  },
  {
    "Classification": "hbase",
    "Properties": {
      "hbase.emr.storageMode": "s3",
      "hbase.emr.readreplica.enabled": "true"
    }
  }
}
```

## Synchronisieren des Lesereplikats beim Hinzufügen von Daten

Da die Read-Replica HBase StoreFiles und Metadaten verwendet, die der primäre Cluster in Amazon S3 schreibt, ist die Read-Replica nur so aktuell wie der Amazon S3 S3-Datenspeicher. Die folgenden Anleitungen können Ihnen dabei helfen, die Zeitverzögerung beim Schreiben von Daten zwischen dem primären Cluster und dem Read Replica zu minimieren.

- Sie sollten nach Möglichkeit Massenladevorgänge für den primären Cluster durchführen. Weitere Informationen finden Sie unter [Massenladung](#) in der Apache-HBase-Dokumentation.
- Möglichst bald nach dem Hinzufügen der Daten sollte eine Auslagerung erfolgen, bei der Speicherdateien in Amazon S3 geschrieben werden. Führen Sie entweder eine manuelle Auslagerung durch oder optimieren Sie die Auslagerungseinstellungen, um Verzögerungen zu minimieren.
- Wenn Verdichtungen möglicherweise automatisch ausgeführt werden, führen Sie eine manuelle Verdichtung durch, um Inkonsistenzen bei der Auslösung von Verdichtungen zu vermeiden.
- Wenn Metadaten geändert wurden (wenn z. B. eine HBase-Regionsaufteilung oder Verdichtungen vorgenommen oder Tabellen hinzugefügt bzw. entfernt werden), führen Sie auf dem Lesereplikat-Cluster den Befehl `refresh_meta` aus.
- Führen Sie den Befehl `refresh_hfiles` auf dem Read Replica-Cluster aus, wenn Datensätze zu einer Tabelle hinzugefügt oder in einer Tabelle geändert werden.



## Persistentes HFile-Tracking

Persistentes HFile-Tracking verwendet eine HBase-Systemtabelle namens `hbase:storefile`, die aufgerufen wird, um die für Lesevorgänge verwendeten HFile-Pfade direkt nachzuverfolgen. Neue HFile-Pfade werden der Tabelle hinzugefügt, sobald zusätzliche Daten zu HBase hinzugefügt werden. Dadurch werden Umbenennungsvorgänge als Commit-Mechanismus in den HBase-Vorgängen für kritische Schreibpfade entfernt und die Wiederherstellungszeit beim Öffnen einer HBase-Region verbessert, indem aus der `hbase:storefile`-Systemtabelle statt aus der Verzeichnisliste des Dateisystems gelesen wird. Dieses Feature ist in Amazon-EMR-Version 6.2.0 und höher standardmäßig aktiviert und erfordert keine manuellen Migrationsschritte.

### Note

Die persistente HFile-Nachverfolgung mithilfe der HBase-Storefile-Systemtabelle unterstützt das Feature zur Replikation von HBase-Regionen nicht. Weitere Informationen zur HBase-Regionsreplikation finden Sie unter [Zeitleistenkonsistente, hochverfügbare Lesevorgänge](#).

## Persistentes HFile-Tracking deaktivieren

Persistentes HFile-Tracking ist ab Amazon EMR Version 6.2.0 standardmäßig aktiviert. Um das persistente HFile-Tracking zu deaktivieren, geben Sie beim Starten eines Clusters die folgende Konfigurationsüberschreibung an:

```
{
  "Classification": "hbase-site",
```

```
"Properties": {
  "hbase.storefile.tracking.persist.enabled":"false",

  "hbase.hstore.engine.class":"org.apache.hadoop.hbase.regionserver.DefaultStoreEngine"
}
}
```

### Note

Bei der Neukonfiguration des Amazon-EMR-Clusters müssen alle Instance-Gruppen aktualisiert werden.

## Manuelles Synchronisieren der Storefile-Tabelle

Die Storefile-Tabelle wird auf dem neuesten Stand gehalten, wenn neue Dateien erstellt werden. Wenn die Storefile-Tabelle jedoch aus irgendeinem Grund nicht mehr mit den Datendateien synchron ist, können die folgenden Befehle verwendet werden, um die Daten manuell zu synchronisieren:

Synchronisieren der Storefile-Tabelle in einer Online-Region:

```
hbase org.apache.hadoop.hbase.client.example.RefreshHFilesClient <table>
```

Synchronisieren der Storefile-Tabelle in einer Offline-Region:

- Entfernen Sie die Storefile-Tabelle znode.

```
echo "ls /hbase/storefile/loaded" | sudo -u hbase hbase zkcli
[<tableName>, hbase:namespace]
# The TableName exists in the list
echo "delete /hbase/storefile/loaded/<tableName>" | sudo -u hbase hbase zkcli
# Delete the Table ZNode
echo "ls /hbase/storefile/loaded" | sudo -u hbase hbase zkcli
[hbase:namespace]
```

- Weisen Sie die Region zu (in 'hbase shell' ausführen).

```
hbase cli> assign '<region name>'
```

- Wenn die Zuweisung fehlschlägt.

```
hbase cli> disable '<table name>'
```

```
hbase cli> enable '<table name>'
```

## Skalierung der Storefile-Tabelle

Die Storefile-Tabelle ist standardmäßig in vier Regionen aufgeteilt. Wenn die Storefile-Tabelle immer noch unter starker Schreiblast steht, kann die Tabelle manuell weiter aufgeteilt werden.

Um eine bestimmte Hot-Region aufzuteilen, verwenden Sie den folgenden Befehl (in 'hbase shell' ausführen).

```
hbase cli> split '<region name>'
```

Um die Tabelle zu teilen, verwenden Sie den folgenden Befehl (in der „hbase-Shell“ ausführen).

```
hbase cli> split 'hbase:storefile'
```

## Betriebliche Überlegungen

Server in der HBase-Region werden verwendet BlockCache , um gelesene Daten im Arbeitsspeicher zu speichern und BucketCache um gelesene Daten auf der lokalen Festplatte zu speichern. Darüber hinaus speichern Regionalserver Datenschreibvorgänge im Arbeitsspeicher und verwenden MemStore Write-Ahead-Protokolle, um Datenschreibvorgänge in HDFS zu speichern, bevor die Daten in StoreFiles HBase in Amazon S3 geschrieben werden. Die Leseleistung Ihres Clusters hängt davon ab, wie oft ein Datensatz vom In-Memory-Cache oder Cache auf dem Datenträger abgerufen werden kann. Ein Cache-Fehler führt dazu, dass der Datensatz aus Amazon S3 gelesen wird, was eine deutlich höhere Latenz und eine höhere Standardabweichung als das Lesen aus HDFS aufweist. StoreFile Darüber hinaus sind die maximalen Anforderungsraten für Amazon S3 geringer als das, was vom lokalen Cache erreicht werden kann. Also kann die Zwischenspeicherung von Daten für leseintensive Workloads sehr wichtig sein. Weitere Informationen zu Amazon-S3-Leistung finden Sie unter [Optimieren der Leistung](#) im Benutzerhandbuch zu Amazon Simple Storage Service.

Um die Leistung zu verbessern, empfehlen wir, so viele Datensätze wie möglich im EC2-Instance-Speicher zwischenspeichern. Da der den EC2-Instance-Speicher des Regionsservers BucketCache verwendet, können Sie einen EC2-Instance-Typ mit einem ausreichenden Instance-Speicher wählen und Amazon EBS-Speicher hinzufügen, um die erforderliche Cache-Größe zu erreichen. Mithilfe der Eigenschaft können Sie auch die BucketCache Größe der

angehängten Instance-Speicher und EBS-Volumes erhöhen. `hbase.bucketcache.size` Die Standardeinstellung lautet 8 192 MB.

Bei Schreibvorgängen können die Häufigkeit von MemStore Flushes und die Anzahl der bei kleineren und größeren Komprimierungen StoreFiles vorhandenen Flushes erheblich zu einer Verlängerung der Antwortzeiten der Server in der Region beitragen. Um eine optimale Leistung zu erzielen, sollten Sie erwägen, den Blockmultiplikator für MemStore Flush und HRegion zu vergrößern. Dadurch erhöht sich die Zeit zwischen größeren Verdichtungen, aber auch die Verzögerung bei der Konsistenz, wenn Sie eine Read-Replica verwenden. Manchmal können Sie eine bessere Leistung erzielen, wenn Sie eine größere Dateiblockgröße verwenden (jedoch weniger als 5 GB), um einen mehrteiligen Amazon-S3-Upload in EMRFS auszulösen. Die Blockgröße von Amazon EMR ist standardmäßig 128 MB. Weitere Informationen finden Sie unter [HDFS-Konfiguration](#). Während dem Benchmarking der Leistung durch Bereinigungen und Verdichtungen sind selten Kunden anzutreffen, die die 1 GB-Blockgröße überschreiten. Darüber hinaus funktionieren HBase-Komprimierungen und Regionserver optimal, wenn weniger komprimiert werden müssen. StoreFiles

Es kann recht lange dauern, bis Tabellen in Amazon S3 gelöscht sind, da große Verzeichnisse umbenannt werden müssen. Erwägen Sie, Tabellen zu deaktivieren, anstatt sie zu entfernen.

Es gibt einen HBase-Bereinigungsprozess, der alte WAL-Dateien und Speicherdateien bereinigt. Bei Amazon-EMR-Version 5.17.0 und höher ist die Bereinigungsfunktion global aktiviert, und die folgenden Konfigurationseigenschaften können verwendet werden, um ein saubereres Verhalten zu steuern.

Konfigurationseigenschaft	Standardwert	Beschreibung
<code>hbase.regionserver.hfilecleaner.large.thread.count</code>	1	Die Anzahl der Threads für die Bereinigung abgelaufener großer HFiles.
<code>hbase.regionserver.hfilecleaner.small.thread.count</code>	1	Die Anzahl der Threads für die Bereinigung abgelaufener kleiner HFiles.
	Setzen Sie diesen Wert auf ein Viertel aller verfügbaren Kerne.	Die Anzahl der Threads zum Durchsuchen der oldWALs Verzeichnisse.

Konfigurationseigenschaft	Standardwert	Beschreibung
<code>hbase.cleaner.scan.dir.concurrent.size</code>		
<code>hbase.oldwals.cleaner.thread.size</code>	2	Die Anzahl der Threads für die Bereinigung der WALs im oldWALS-Verzeichnis.

Bei Amazon EMR 5.17.0 und früher kann sich der Bereinigungsverfahren bei hohen Workloads auf die Abfrageleistung auswirken. Deshalb wird empfohlen, die Bereinigung nur außerhalb von Spitzenzeiten durchzuführen. Die Bereinigung umfasst die folgenden HBase-Shell-Befehle:

- `cleaner_chore_enabled` fragt ab, ob die Bereinigungsfunktion aktiviert ist.
- `cleaner_chore_run` führt die Bereinigungsfunktion manuell aus, um Dateien zu entfernen.
- `cleaner_chore_switch` aktiviert oder deaktiviert die Bereinigungsfunktion und gibt den vorherigen Zustand der Bereinigungsfunktion zurück. Durch `cleaner_chore_switch true` wird die Bereinigungsfunktion beispielsweise aktiviert.

## Eigenschaften für die Leistungsoptimierung von HBase auf Amazon S3

Die folgenden Parameter können so angepasst werden, dass die Leistung Ihres Workloads bei Verwendung von HBase auf Amazon S3 optimiert wird.

Konfigurationseigenschaft	Standardwert	Beschreibung
<code>hbase.bucketcache.size</code>	8,192	Die Menge an Festplattenspeicher in MB, die auf Regionsservern, Amazon EC2 EC2-Instance-Speichern und EBS-Volumes für BucketCache die Speicherung reserviert ist. Die Einstellung gilt für alle Regionsserver-Instances. Größere BucketCache Größen

Konfigurationseigenschaft	Standardwert	Beschreibung
		entsprechen in der Regel einer verbesserten Leistung
<code>hbase.hregion.memstore.flush.size</code>	134217728	Das Datenlimit in Byte, bei dem eine MemStore-Auslagerung nach Amazon S3 ausgelöst wird.
<code>hbase.hregion.memstore.block.multiplier</code>	4	Ein Multiplikator, der die MemStore Obergrenze bestimmt, ab der Updates blockiert werden. Wenn die MemStore <code>hbase.hregion.memstore.flush.size</code> Überschreitungen mit diesem Wert multipliziert werden, werden Updates blockiert. MemStore Es kann vorkommen, dass Updates durch Leeren und Komprimieren entsperrt werden.
<code>hbase.hstore.blockingStoreFiles</code>	10	Die maximale Anzahl StoreFiles davon kann in einem Store existieren, bevor Updates blockiert werden.
<code>hbase.hregion.max.filesize</code>	10737418240	Die maximale Größe einer Region, bevor die Region aufgeteilt wird.

## Herunterfahren und Wiederherstellen eines Clusters ohne Datenverlust

Um einen Amazon EMR-Cluster herunterzufahren, ohne Daten zu verlieren, die nicht in Amazon S3 geschrieben wurden, sollten Sie Ihren MemStore Cache auf Amazon S3 leeren, um neue Speicherdateien zu schreiben. Zunächst müssen Sie alle Tabellen deaktivieren. Die folgende

Schrittkonfiguration kann verwendet werden, wenn Sie dem Cluster einen Schritt hinzufügen. Weitere Informationen finden Sie unter [Arbeiten mit Schritten unter Verwendung der AWS CLI und Konsole](#) im Verwaltungshandbuch für Amazon EMR.

```
Name="Disable all tables",Jar="command-runner.jar",Args=["/bin/bash","/usr/lib/hbase/bin/disable_all_tables.sh"]
```

Alternativ können Sie auch den folgenden Bash-Befehl direkt ausführen.

```
bash /usr/lib/hbase/bin/disable_all_tables.sh
```

Nachdem Sie alle Tabellen deaktiviert haben, leeren Sie die `hbase:meta`-Tabelle mit der HBase-Shell und dem folgenden Befehl.

```
flush 'hbase:meta'
```

Anschließend können Sie ein auf dem Amazon EMR-Cluster bereitgestelltes Shell-Skript ausführen, um den MemStore Cache zu leeren. Sie können es als einen Schritt hinzufügen oder direkt auf dem Cluster AWS CLI ausführen. Das Skript deaktiviert alle HBase-Tabellen, was dazu führt, dass der Server MemStore in jeder Region zu Amazon S3 flusht. Wenn das Skript erfolgreich abgeschlossen wurde, verbleiben die Daten in Amazon S3 und der Cluster kann beendet werden.

Um einen Cluster mit denselben HBase-Daten neu zu starten, geben Sie denselben Amazon S3 S3-Standort wie der vorherige Cluster entweder in der AWS Management Console oder mithilfe der `hbase.rootdir` Konfigurationseigenschaft an.

## Write-Ahead-Protokolle (WAL) für Amazon EMR

Mit Amazon EMR 6.15 und höher können Sie Ihre Write-Ahead-Protokolle (WAL) von Apache HBase in Amazon EMR WAL schreiben. Wenn Sie bei niedrigeren Amazon-EMR-Versionen einen Cluster mit der Option HBase auf Amazon S3 erstellen, ist WAL die einzige Apache-HBase-Komponente, die auf der lokalen Festplatte für Cluster gespeichert wird, und Sie können andere Komponenten wie das Stammverzeichnis, Dateien speichern (HFiles Tabellenmetadaten und Daten auf Amazon S3 speichern).

Sie können Amazon EMR WAL verwenden, um Daten wiederherzustellen, die nicht in Amazon S3 geleert wurden. Um Ihre HBase-Cluster vollständig zu sichern, melden Sie sich für die Verwendung des Amazon-EMR-WAL-Service an. Im Hintergrund `RegionServers` schreibt Ihre HBase-Write-Ahead-Protokolle (WAL) in das WAL für Amazon EMR.



Für den Fall, dass Ihr Cluster oder die AZ fehlerhaft oder nicht verfügbar wird, können Sie einen neuen Cluster erstellen, ihn auf dasselbe S3-Stammverzeichnis und denselben Amazon-EMR-WAL-Workspace verweisen und die Daten innerhalb weniger Minuten automatisch in WAL wiederherstellen. Weitere Informationen finden Sie unter [Wiederherstellen aus Amazon EMR WAL](#).

### Note

Amazon EMR speichert Ihr Write-Ahead-Protokoll und seine Daten 30 Tage ab dem Zeitpunkt, an dem Sie Ihren Cluster erstellen. Nach 30 Tagen löscht Amazon EMR automatisch Ihr Amazon EMR WAL und seine Daten. Wenn Sie jedoch einen neuen WAL-fähigen Cluster aus demselben S3-Stammverzeichnis starten, können Sie die Verwendung Ihres WAL um 30 Tage ab der Startzeit des neuen Clusters verlängern. Amazon EMR bereinigt nach den ersten 30 Tagen weiterhin alle WAL-Daten aus dem ersten Cluster. Weitere Informationen finden Sie unter [Wiederherstellen aus Amazon EMR WAL](#).

In den folgenden Abschnitten wird beschrieben, wie Sie Amazon EMR WAL mit Ihrem HBase-fähigen EMR-Cluster einrichten und verwenden.

### Themen

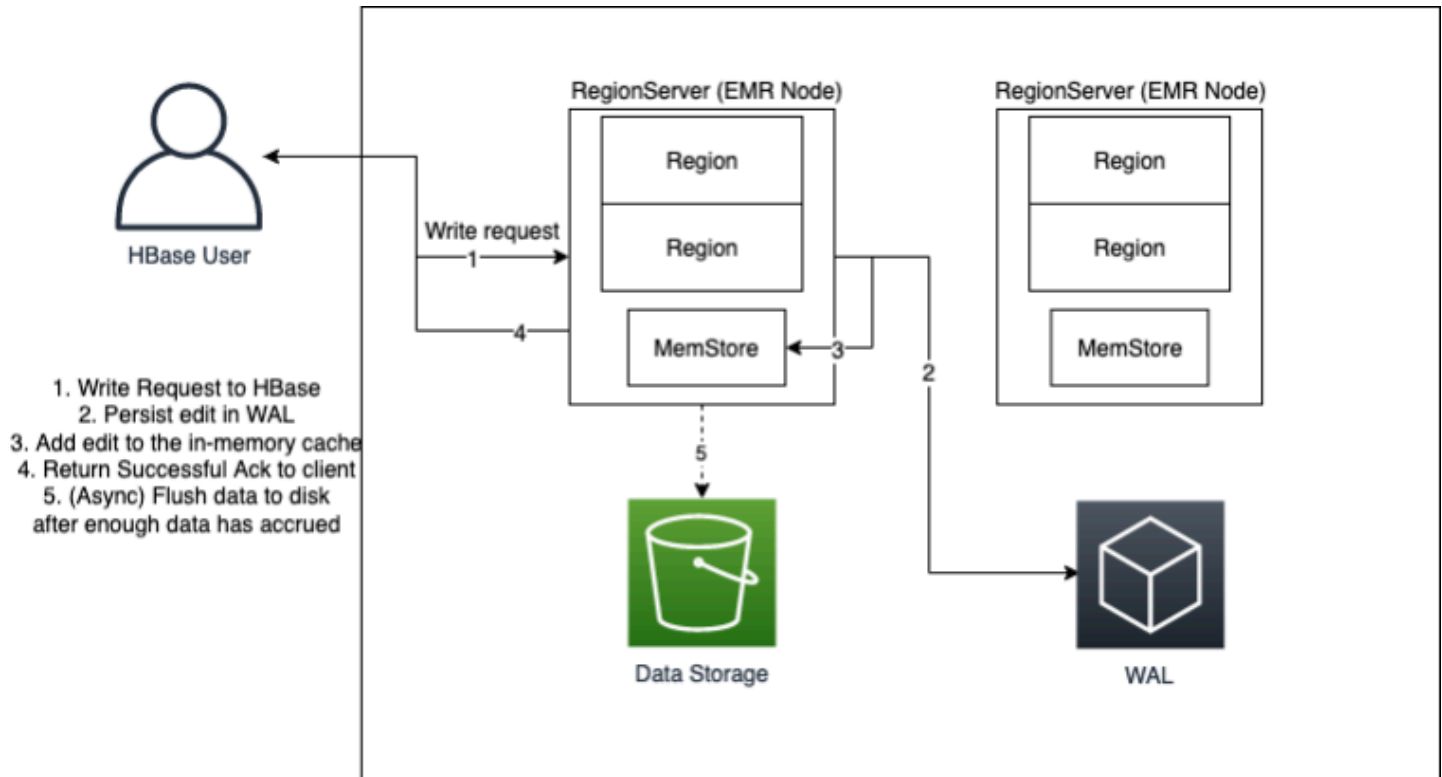
- [Amazon-EMR-WAL-Workspaces](#)
- [Erforderliche Berechtigungen für Amazon EMR WAL](#)
- [Aktivieren von Amazon EMR WAL](#)
- [Wiederherstellen aus Amazon EMR WAL](#)
- [Verwenden von Sicherheitskonfigurationen mit Amazon EMR WAL](#)
- [Zugriff auf Amazon EMR WAL über AWS PrivateLink](#)
- [Grundlegendes zu Amazon-EMR-WAL-Preisen und -Metriken](#)
- [Markieren von WAL-Workspaces](#)
- [Überlegungen und Regionen für Amazon EMR WAL](#)
- [Amazon-EMR-WAL-\(EMRWAL\)-CLI-Referenz](#)

## Amazon-EMR-WAL-Workspaces

Amazon EMR WAL fügt das Konzept der WAL-Workspaces hinzu. Ein WAL-Workspace ist ein logischer Container von WALs . Jedes Write-Ahead-Protokoll in Amazon EMR WAL wird durch

einen WAL-Workspace gekapselt. Ein EMR-Cluster schreibt WALs in genau einen WAL-Workspace, den Sie beim Clusterstart konfigurieren, oder in den , defaultWALworkspace wenn Sie keinen Workspace angeben. WAL-Workspaces stehen in keinem Zusammenhang mit vorhandenen HBase-Terminologien wie Namespaces.

Sie können WAL-Workspaces verwenden, um Amazon-EMR-WAL-IAM-Berechtigungen so einzuschränken, dass sie nur die Workspaces enthalten, auf die der Cluster zugreifen muss. Sie können Ihren WAL-Workspace auch für die Tag-basierte Zugriffskontrolle markieren. Weitere Informationen zum Markieren finden Sie unter [Markieren von WAL-Workspaces](#).



## Erforderliche Berechtigungen für Amazon EMR WAL

Damit Ihr Cluster eine Verbindung zu Amazon EMR WAL herstellen kann, benötigt das Instance-Profil für den Cluster bestimmte IAM-Berechtigungen:

- Amazon EMR WAL verwendet die [AWSServiceRoleForEMRWAL](#) serviceverknüpfte Rolle, um einen Cluster-Status abzurufen. Amazon EMR erstellt diese serviceverknüpfte Rolle automatisch, wenn Sie einen WAL-Workspace erstellen, oder HBase erstellt die serviceverknüpfte Rolle, wenn Sie einen Workspace für Amazon EMR WAL konfigurieren und die serviceverknüpfte Rolle noch nicht existiert.

Bevor Sie Amazon EMR WAL für einen Cluster aktivieren können, müssen Sie die Berechtigungen konfigurieren, um die automatische Erstellung der `AWSServiceRoleForEMRWAL` serviceverknüpften Rolle zu ermöglichen. Weitere Informationen und eine Beispielanweisung, die diese Funktion hinzufügt, finden Sie unter [Verwenden von serviceverknüpften Rollen für die Write-Ahead-Protokollierung](#).

- Da Amazon EMR WAL HBase Write Ahead Log (WAL) verwendet, müssen Ihre Cluster HBase WAL verwenden. Im Folgenden sind die IAM-Mindestberechtigungen aufgeführt, die Sie zum Ausführen von HBase benötigen. Fügen Sie diese der Berechtigungsrichtlinie für Ihr Instance-Profil hinzu:

```
emrwal:DeleteWal
emrwal:CreateWal
emrwal:CreateWorkspace
emrwal:AppendEdit
emrwal:ReplayEdits
emrwal:GetCurrentWalTime
emrwal:CompleteWalFlush
```

#### Note

Wenn Sie die Berechtigungen für Amazon EMR WAL auf den Minimalersatz beschränken, verfügen einige [EMRWAL CLI](#)-Befehle nicht über die erforderlichen Berechtigungen zum Ausführen.

## Aktivieren von Amazon EMR WAL

Gehen Sie wie folgt vor, um das Schreiben in das Amazon EMR WAL zu aktivieren, wenn Sie einen Cluster mit `erstellenAWS Command Line Interface`.

#### Note

Sie können Amazon EMR WAL nicht für einen Cluster aktivieren, der bereits ausgeführt wird, und Sie können nicht zwei Cluster mit demselben S3-Stammverzeichnis starten. Weitere Informationen finden Sie unter [Überlegungen und Regionen für Amazon EMR WAL](#).

1. Bevor Sie einen Amazon-EMR-WAL-fähigen Cluster erstellen können, müssen Sie dem Instance-Profil, das Sie mit Ihrem Cluster verwenden möchten, die erforderlichen Berechtigungen hinzufügen. Weitere Informationen finden Sie unter [Erforderliche Berechtigungen für Amazon EMR WAL](#).
2. Erstellen Sie einen Cluster über die AWS CLI. Verwenden Sie die `--configurations` Option, um ein JSON-Konfigurationsobjekt bereitzustellen, das die `hbase.emr.wal.enabled` Eigenschaft angibt, wie im folgenden Beispiel gezeigt.
  - Geben Sie den Speichermodus und den Speicherort des Stammverzeichnisses in Amazon S3 an. Der von Ihnen angegebene Amazon S3-Speicherort sollte sich in derselben Region wie Ihr EMR-Cluster befinden, aber nur ein aktiver Cluster kann gleichzeitig dasselbe HBase-Stammverzeichnis in S3 verwenden.
  - Erstellen Sie Ihren Cluster mit der Instance-Gruppenkonfiguration. Sie können Amazon EMR WAL nicht mit der Konfiguration der Instance-Flotten verwenden. Weitere Informationen zum Erstellen von Clustern mit Instance-Gruppen finden Sie unter [Konfigurieren einheitlicher Instance-Gruppen](#) im Verwaltungshandbuch für Amazon EMR.
  - Schritte zur Konsole zum Erstellen eines Clusters und ein detailliertes `create-cluster` Beispiel, das verwendet AWS CLI, finden Sie unter [Erstellen eines Clusters mit HBase](#).
3. Um WAL für den neuen Cluster zu aktivieren, setzen Sie die `-hbase.emr.wal.enabled` Eigenschaft auf `true`. Der folgende Befehl enthält einen JSON-Ausschnitt mit einem Beispielkonfigurationsobjekt.

```
aws emr create-cluster --name "hbasewal" --release-label emr-6.x.y \  
--applications Name=HBase --use-default-roles --ec2-attributes KeyName=myKey \  
--instance-type m6i.xlarge --instance-count 1 --configurations hbase.json  
$cat hbase.json  
[  
  {  
    "Classification": "hbase-site",  
    "Properties": {  
      "hbase.rootdir": "s3://MyBucket/MyHBaseStore"  
    }  
  },  
  {  
    "Classification": "hbase",  
    "Properties": {  
      "hbase.emr.storageMode": "s3",  
      "hbase.emr.wal.enabled": "true"  
    }  
  }  
]
```

```
}  
]
```

Wenn HBase auf dem neu erstellten Cluster online ist, schreibt HBase automatisch WAL-Daten in das Amazon-EMR-WAL und verwendet das Amazon-EMR-WAL für Wiederherstellungszwecke.

Example 1: Erstellen eines EMR-Clusters, der Amazon EMR WAL verwendet

```
[  
  {  
    "Classification": "hbase-site",  
    "Properties": {  
      "hbase.rootdir": "s3://MyBucket/MyHBaseStore"  
    }  
  },  
  {  
    "Classification": "hbase",  
    "Properties": {  
      "hbase.emr.storageMode": "s3",  
      "hbase.emr.wal.enabled": "true"  
    }  
  }  
]
```

Example 2: Erstellen eines EMR-Clusters mit einem benutzerdefinierten WAL-Workspace

```
[  
  {  
    "Classification": "hbase-site",  
    "Properties": {  
      "hbase.rootdir": "s3://MyBucket/MyHBaseStore",  
      "emr.wal.workspace": "customWorkspaceName"  
    }  
  },  
  {  
    "Classification": "hbase",  
    "Properties": {  
      "hbase.emr.storageMode": "s3",  
      "hbase.emr.wal.enabled": "true"  
    }  
  }  
]
```

]

## Wiederherstellen aus Amazon EMR WAL

Da das Amazon EMR WAL für Ihren ursprünglichen Cluster 30 Tage lang aufbewahrt wird, können Sie das WAL für einen neu erstellten Cluster innerhalb dieses 30-Tage-Zeitraums wiederherstellen und wiederverwenden. Wenn Sie einen neuen Cluster aus demselben S3-Stammverzeichnis starten, wird die 30-Tage-Uhr ab der Startzeit des neuen Clusters neu gestartet, solange der vorherige 30-Tage-Zeitraum nicht abgelaufen ist.

Gehen Sie wie folgt vor, um ein vorhandenes WAL mit einem neuen Cluster wiederherzustellen. Bei diesem Prozess wird davon ausgegangen, dass Sie Ihren ursprünglichen Cluster mit aktiviertem Amazon EMR WAL erstellt haben.

1. Erstellen Sie innerhalb von 30 Tagen nach dem Erstellen eines WAL-fähigen Clusters einen neuen Cluster in derselben AWS-Region wie der ursprüngliche Cluster. Der neue Cluster kann sich in derselben AZ oder in einer anderen AZ innerhalb derselben Region befinden, in der der ursprüngliche Cluster erstellt wurde.

Konfigurieren Sie die Objekteigenschaften, um den Speichermodus und den Speicherort des Stammverzeichnisses in Amazon S3 anzugeben. Der von Ihnen angegebene Amazon S3-Speicherort sollte sich in derselben Region wie Ihr EMR-Cluster befinden, aber nur ein aktiver Cluster kann gleichzeitig dasselbe HBase-Stammverzeichnis in S3 verwenden.

Schritte zur Konsole zum Erstellen eines Clusters und ein detailliertes `create-cluster` Beispiel, das verwendet AWS CLI, finden Sie unter [Erstellen eines Clusters mit HBase](#).

2. Um das vorhandene Amazon EMR WAL für den neuen Cluster zu verwenden, setzen Sie die `hbase.emr.wal.enabled` Eigenschaft auf `true`. Der folgende JSON-Ausschnitt zeigt ein Beispiel für ein Konfigurationsobjekt.

```
[
  {
    "Classification": "hbase-site",
    "Properties": {
      "hbase.rootdir": "s3://MyBucket/MyHBaseStore"
    }
  },
  {
    "Classification": "hbase",
```

```
    "Properties": {
      "hbase.emr.storageMode": "s3",
      "hbase.emr.wal.enabled": "true"
    }
  }
]
```

## Verwenden von Sicherheitskonfigurationen mit Amazon EMR WAL

Amazon EMR verschlüsselt automatisch sowohl Daten während der Übertragung zwischen Ihrem Cluster und dem Amazon-EMR-WAL-Service als auch die Daten im Ruhezustand in Amazon EMR WAL. Weitere Informationen finden Sie unter [Verschlüsselung im Ruhezustand für Amazon EMR WAL](#). Sie können auch eine Sicherheitskonfiguration verwenden, um Ihre eigenen Schlüssel vom AWS Key Management Service (KMS)-Service mitzubringen und die Daten zu verschlüsseln, die Sie in Amazon EMR WAL speichern.

Verwenden Sie eine der folgenden Methoden, um eine Sicherheitskonfiguration auszuwählen, wenn Sie einen Cluster erstellen:

### Console

AWS Management Console Geben Sie in der die Konfiguration unter Sicherheitskonfiguration und EC2-Schlüsselpaar an.

#### Security configuration and EC2 key pair - *optional* [Info](#)

##### Security configuration

Select your cluster encryption, authentication, authorization, and instance metadata service settings.



### CLI

Legen Sie in der den `--security-configuration` Parameter fest AWS CLI, wenn Sie den Befehl [create-cluster](#) verwenden.

Weitere Informationen finden Sie unter [Verschlüsselung im Ruhezustand für Amazon EMR WAL](#) und [Verwenden von Sicherheitskonfigurationen zum Einrichten der Clustersicherheit im Verwaltungshandbuch für Amazon EMR](#).

Weitere sicherheitsrelevante Informationen zu WAL finden Sie unter [Verwenden von serviceverknüpften Rollen für die Write-Ahead-Protokollierung](#).

## Zugriff auf Amazon EMR WAL über AWS PrivateLink

Wenn Sie Ihre Verbindung innerhalb des AWS Netzwerks behalten möchten, bietet Amazon EMR WAL AWS PrivateLink Unterstützung. Um einzurichten AWS PrivateLink, verwenden Sie die AWS Management Console oder AWS Command Line Interface (AWS CLI), um einen Schnittstellen-VPC-Endpunkt zu erstellen, der eine Verbindung zu Amazon EMR WAL herstellt. Weitere Informationen finden Sie unter [Zugriff auf einen AWS-Service über einen Schnittstellen-VPC-Endpunkt](#) im AWS PrivateLink-Leitfaden.

Die grundlegenden Schritte sind wie folgt:

1. Verwenden Sie die Amazon-VPC-Konsole, um [einen VPC-Endpunkt zu erstellen](#). Wählen Sie Endpunkte und dann Endpunkt erstellen aus.
2. Behalten Sie die Kategorie Service als AWS Services bei.
3. Geben Sie in der Suchleiste für das Services-Bedienfeld ein **emrwal** und wählen Sie dann den Service mit der Bezeichnung `com.amazonaws.region.emrwal.prod`.
4. Wählen Sie Ihre VPC aus und speichern Sie den Endpunkt. Stellen Sie sicher, dass Sie dieselben Sicherheitsgruppen an den VPC-Endpunkt anfügen, den Sie dem EMR-Cluster anfügen.
5. Wenn Sie möchten, können Sie jetzt private DNS-Hostnamen für Ihren neuen Endpunkt aktivieren. Stellen Sie DNS-Hostnamen aktivieren und DNS-Unterstützung aktivieren auf `true` für Ihre VPC ein. Wählen Sie dann Ihre Endpunkt-ID aus, wählen Sie im Menü Aktionen die Option VPC-Einstellungen bearbeiten und aktivieren Sie private DNS-Namen.
  - Die privaten DNS-Hostnamen für den Endpunkt folgen dem Format `prod.emrwal.region.amazonaws.com`.
  - Wenn Sie keine privaten DNS-Hostnamen aktivieren, stellt Amazon VPC einen DNS-Endpunktnamen für Sie im Format `bereitendpointID.prod.emrwal.region.vpce.amazonaws.com`.
6. Um Ihren AWS PrivateLink Endpunkt zu verwenden, ändern Sie die `emr.wal.client.endpoint` Konfiguration, wenn Sie Ihren [Amazon-EMR-WAL-fähigen Cluster](#) erstellen, wie im folgenden Beispiel gezeigt:

```
[
```



```

{
  "Classification": "hbase-site",
  "Properties": {
    "hbase.rootdir": "s3://MyBucket/MyHBaseStore",
    "emr.wal.workspace": "customWorkspaceName",
    "emr.wal.client.endpoint": "https://prod.emrwal.region.amazonaws.com"
  }
},
{
  "Classification": "hbase",
  "Properties": {
    "hbase.emr.storageMode": "s3",
    "hbase.emr.wal.enabled": "true"
  }
}
]

```

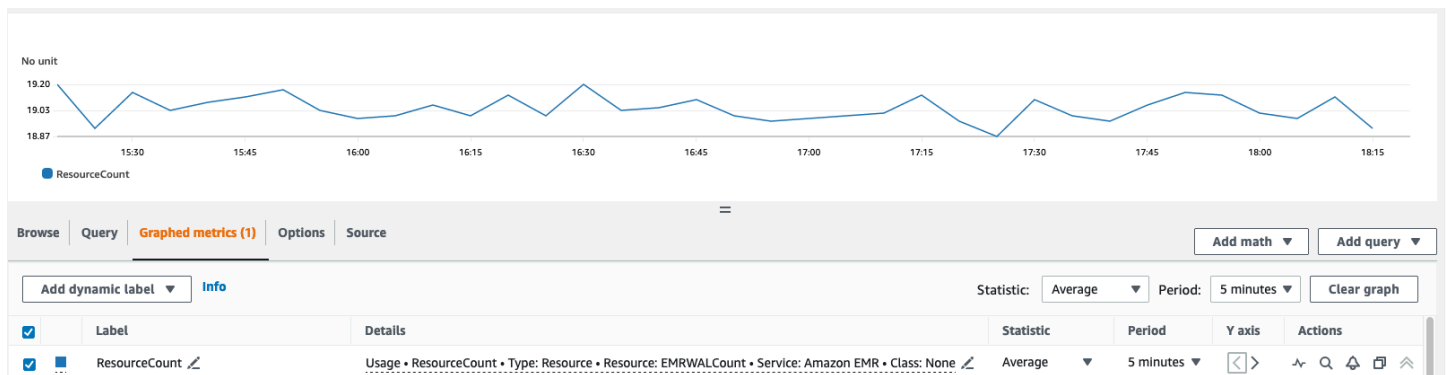
Sie können auch die VPCE-Richtlinie verwenden, um den Zugriff auf die Amazon-EMR-WAL-APIs zu erlauben oder einzuschränken. Weitere Informationen finden Sie unter [Steuern des Zugriffs auf VPC-Endpunkte mithilfe von Endpunktrichtlinien](#) im AWS PrivateLink -Handbuch.

## Grundlegendes zu Amazon-EMR-WAL-Preisen und -Metriken

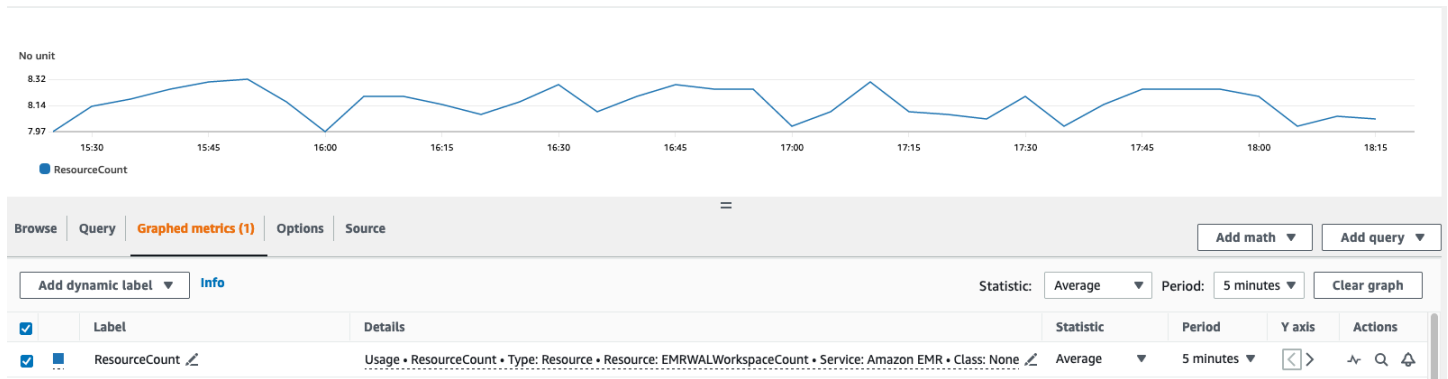
Fakturierungseinheit für Kernfunktionen	Details
EMR-WAL-Read-GiB	API calls to read data from your table are billed as ReadRequestGiB. This includes <a href="#">Get and Scan</a> operations. Reads are charged based on the sizes of the read items. Amazon EMR bills at a minimum of 1 byte. For example, if you read a 1234.12 bytes item, you're charged for 1235 bytes. Reads are aggregated every hour for billing and shown as GiBs.
EMR-WAL-Write-GiB	API calls to write data from your table are billed as Write-GiB. This includes <a href="#">Put</a> operations. Writes are charged based on the sizes of the written items. Amazon EMR bills at a minimum of 1 byte. For example, if you write a 1234.12 bytes item, you're charged for 1235

<b>Fakturierungseinheit für Kernfunktionen</b>	<b>Details</b>
	bytes. Writes are aggregated every hour for billing and shown as GiBs.
<b>EMR-WAL-WALHours</b>	<p>Die Anzahl der WALs, die Sie im Service speichern, wird als <math>\text{abgerechnetEMR-WAL-WALHours}</math> . Amazon EMR erstellt ein WAL pro HBase-Region. Wenn Sie beispielsweise 20 HBase-Tabellen einschließlich Systemtabellen erstellen und jede Tabelle zwei HBase-Regionen hat, verwenden Sie 28 800 WAL-Stunden, berechnet als:</p> <div style="border: 1px solid #ccc; border-radius: 10px; padding: 10px; margin: 10px 0;"> <p style="margin: 0;">                         20 tables                          x 2 Regions per table                          x 1 WAL per Region                          x 30 days                          x 24 hours                          -----                          28,800 EMR-WAL-WALHours                     </p> </div>

**BeispielEMRWALCount:**



**BeispielEMRWALWorkspaceCount:**



## Markieren von WAL-Workspaces

Sie können einem Workspace Tags hinzufügen, wenn Sie einen neuen Workspace erstellen, und Sie können Tags aus einem aktiven Workspace für einen laufenden Cluster hinzufügen, entfernen oder auflisten. Sie können die einzelnen Ressourcen im Workspace nicht markieren und vorhandene Tags nicht aktualisieren. Entfernen Sie stattdessen unerwünschte Tags aus dem Workspace und ersetzen Sie sie.

Sie können Workspaces über die markierenEMRWAL CLI. Eine Liste der EMRWAL CLI Befehle zum Markieren von Workspaces finden Sie unter [Amazon-EMR-WAL-\(EMRWAL\)-CLI-Referenz](#).

Die folgende IAM-Beispielrichtlinie veranschaulicht ein Szenario, das Workspace-CRUDL-Operationen nur mit dem richtigen Tagging-Schlüssel `resource_tag_allow_test_key` und -Wert zulässt `resource_tag_allow_test_value`:

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Action": [
        "EMRWAL:*"
      ],
      "Effect": "Allow",
      "Resource": [
        "*",
        "*"
      ],
      "Condition": {
        "StringEquals": {
          "aws:ResourceTag/resource_tag_allow_test_key": [
            "resource_tag_allow_test_value"
          ]
        }
      }
    }
  ]
}
```

```

    ]
  }
}
]
}

```

Um zu überprüfen, ob das Tag jetzt für Workspace-Operationen erforderlich ist, verwenden Sie die , [Amazon-EMR-WAL-\(EMRWAL\)-CLI-Referenz](#) um den [listTagsForResource](#) Befehl auf tagAllowResourceTag für den Workspace mit dem gewünschten Ressourcen-Tag aufzurufen. Wenn Sie die Bedingung korrekt konfiguriert haben, ist der Befehl erfolgreich.

```

emrwal listTagsForResource -r us-east-1 -arn arn:aws:emrwal:us-east-1:arn:workspace/
tagAllowResourceTag
Tag(Key=resource_tag_allow_test_key, Value=resource_tag_allow_test_value)

```

## Überlegungen und Regionen für Amazon EMR WAL

### Überlegungen zu Amazon EMR WAL

In der folgenden Liste werden wichtige Überlegungen und Einschränkungen von Amazon EMR WAL beschrieben:

- Amazon EMR WAL ist für die Verwendung mit Amazon-EMR-Versionen 6.15.0 und höher verfügbar.
- Amazon EMR WAL ist ein kostenpflichtiger Opt-in-Service. Sie zahlen für das, was Sie tatsächlich nutzen: Lesevorgänge, Schreibvorgänge und Datenspeicherung. Weitere Informationen finden Sie unter [Grundlegendes zu Amazon-EMR-WAL-Preisen und -Metriken](#) und auf der Seite [Amazon-EMR-Preise](#).
- Amazon EMR WAL verwendet HBase Write Ahead Log (WAL). Um Amazon EMR WAL verwenden zu können, müssen Ihre Cluster HBase WAL verwenden.
- Um Amazon EMR WAL beim Erstellen eines Clusters zu aktivieren, müssen Sie über die erforderlichen Rollenberechtigungen verfügen. Weitere Informationen finden Sie unter [Verwenden von serviceverknüpften Rollen für die Write-Ahead-Protokollierung](#).
- Sie müssen Amazon EMR WAL aktivieren, wenn Sie den Cluster mit der AWS Management Console, AWS CLI oder API erstellen, und Sie müssen die Instance-Gruppenkonfiguration verwenden. Sie können Amazon EMR WAL nicht in einem laufenden Cluster aktivieren, wenn Sie den Cluster nicht mit Amazon EMR WAL erstellt haben. Sie können die `hbase-site`

Konfigurationen auch nicht bearbeiten, um Amazon EMR WAL in einem laufenden Cluster zu aktivieren.

- Sie können Amazon EMR WAL nur auf Clustern aktivieren, die Amazon S3 für das Stammverzeichnis verwenden.
- Sie können nicht mehrere aktive Cluster im selben HBase-Stammverzeichnis in Amazon S3 haben.
- Sie können Amazon EMR WAL nicht auf Lesereplikat-Clustern aktivieren.
- WAL wird über Availability Zones innerhalb des verwalteten Services repliziert.
- WAL führt den Cluster aus und bleibt für den nächsten Cluster verfügbar.
- Sie können Amazon EMR WAL nicht während des Starts oder wenn Ihr Cluster betriebsbereit ist (in einem laufenden Zustand) deaktivieren.
- Informationen zu WAL- und Workspace-Limits finden Sie unter [Amazon-EMR-Endpunkte und -Kontingente](#).

## Regionale Verfügbarkeit für Amazon EMR WAL

Der Amazon-EMR-WAL-Service ist in den folgenden verfügbarAWS-Regionen:

- `ap-northeast-1` – Asien-Pazifik (Tokio)
- `ap-southeast-1` – Asien-Pazifik (Singapur)
- `ap-south-1` – Asien-Pazifik (Mumbai)
- `ap-southeast-2` – Asien-Pazifik (Sydney)
- `eu-central-1` – Europa (Frankfurt)
- `eu-north-1` – Europa (Stockholm)
- `eu-west-1` – Europa (Irland)
- `us-east-1` – USA Ost (Nord-Virginia)
- `us-east-2` – USA Ost (Ohio)
- `us-west-2` – USA West (Oregon)

## Amazon-EMR-WAL-(EMRWAL)-CLI-Referenz

Die `-EMRWAL`Befehlszeilenschnittstelle (EMRWAL CLI) ist ein einheitliches Tool zur Verwaltung Ihres Write-Ahead-Protokolls (WAL) für Amazon EMR. Die wird mit EMR-Clustern EMRWAL CLI geliefert, wenn Sie WAL zum Zeitpunkt der Erstellung eines Clusters aktivieren. Weitere

Informationen zum Aktivieren von WAL finden Sie unter [Write-Ahead-Protokolle \(WAL\) für Amazon EMR](#).

Die EMRWAL CLI enthält die folgenden Befehle:

Themen

- [createWorkspace](#)
- [deleteWal](#)
- [deleteWorkspace](#)
- [listTagsForResource](#)
- [listWals](#)
- [listWorkspaces](#)
- [tagResource](#)
- [untagResource](#)

## createWorkspace

Der createWorkspace Befehl erstellt einen neuen Amazon-EMR-WAL-Workspace.

Verwendung:

```
emrwal createWorkspace [-tags <tags>] [-e {endpoint}] [-r {Region}] -w {workspacename} [-h]
```

Beispiel:

```
aws emr createWorkspace -w examplews
```

## deleteWal

Der deleteWals Befehl löscht das von Ihnen angegebene Amazon EMR WAL.

Verwendung:

```
emrwal deleteWal [-e {endpoint}] [-r {Region}] [-w {workspacename}] [-p <tablePrefix>] [-n <walName>] [-N <fullName>] [-R] [-m] [-h]
```

Beispiel:

```
emrwal deleteWal -w examples -p hbaseable -n examplewal
```

## deleteWorkspace

Der deleteWorkspace Befehl löscht den von Ihnen angegebenen Amazon-EMR-WAL-Workspace.

Verwendung:

```
emrwal deleteWorkspace [-e {endpoint}] [-r {Region}] -w {workspacename} [-h]
```

Beispiel:

```
emrwal deleteWorkspace -w examples
```

## listTagsForResource

Der listTagsForResource Befehl listet alle Schlüssel-Wert-Paar-Tags für den von Ihnen angegebenen Amazon-EMR-WAL-Workspace auf.

Verwendung:

```
emrwal listTagsForResource -arn {resource-arn} [-e {endpoint}] [-r {Region}] [-h]
```

Beispiel:

```
emrwal listTagsForResource -arn arn:aws:emrwal::1234567891234:workspace/examples
```

## listWals

Der listWals Befehl listet alle Amazon-EMR-WALs in dem von Ihnen angegebenen Workspace auf.

Verwendung:

```
emrwal listWals [-nextToken {token-string}] [-pageSize {integer}] [-e {endpoint}] [-r {Region}] [-w {workspacename}] [-p <tablePrefix>] [-M {integer}] [-h]
```

Beispiel:

```
emrwal listWals -w examplews
```

## listWorkspaces

Der `listWorkspaces` Befehl listet alle Amazon-EMR-WAL-Workspaces auf, die Ihnen zur Verfügung stehen.

Verwendung:

```
emrwal listWorkspaces [-nextToken {token-string}] [-pageSize {integer}] [-e {endpoint}]  
[-r {Region}] [-M {integer}] [-h]
```

Beispiel:

```
emrwal listWorkspaces
```

## tagResource

Der `tagResource` Befehl weist dem von Ihnen angegebenen Amazon-EMR-WAL-Workspace ein oder mehrere Schlüssel-Wert-Paar-Tags zu.

Verwendung:

```
emrwal tagResource -arn {resource-arn} -tags <tags> [-e {endpoint}] [-r {Region}] [-h]
```

Beispiel:

```
emrwal tagResource -arn arn:aws:emrwal::1234567891234:workspace/examplews -  
tags tag_key=tag_value
```

## untagResource

Der `untagResource` Befehl hebt die Zuweisung eines oder mehrerer Schlüssel-Wert-Paar-Tags zu dem von Ihnen angegebenen Amazon-EMR-WAL-Workspace auf.

Verwendung:



```
emrwal untagResource -arn {resource-arn} -tagKeys <tagKeys> [-e {endpoint}] [-r {Region}] [-h]
```

Beispiel:

```
emrwal untagResource -arn arn:aws:emrwal::1234567891234:workspace/examples -tagKeys tag_key
```

## Verwenden der HBase-Shell

Nachdem Sie einen HBase-Cluster erstellt haben, besteht der nächste Schritt darin, eine Verbindung zu HBase herzustellen, damit Sie mit dem Lesen und Schreiben von Daten beginnen können (auf einem Read Replica-Cluster können keine Daten geschrieben werden). Sie können die [HBase-Shell](#) verwenden, um Befehle zu testen.

So öffnen Sie die HBase-Shell

1. Verwenden Sie SSH, um eine Verbindung zum Hauptserver im HBase-Cluster herzustellen. Informationen zum Herstellen einer Connect dem primären Amazon EMR-Knoten mithilfe von SSH finden Sie unter [Verbindung zum primären Knoten mithilfe von SSH](#) herstellen im Amazon EMR Management Guide.
2. Führen Sie `hbase shell`. Die HBase-Shell wird mit einer Eingabeaufforderung geöffnet, ähnlich wie im folgenden Beispiel.

```
hbase(main):001:0>
```

Sie können HBase-Shell-Befehle von der Eingabeaufforderung ausführen. Um weitere Informationen zu den Shell-Befehlen und deren Aufruf zu erhalten, geben Sie "help" an der HBase-Eingabeaufforderung ein und drücken Sie die Eingabetaste.

## Erstellen einer Tabelle

Mit dem folgenden Befehl wird eine Tabelle namens „t1“ erstellt, die eine Spaltengruppe mit dem Namen „f1“ enthält.

```
hbase(main):001:0>create 't1', 'f1'
```

## Eingeben eines Werts

Mit dem folgenden Befehl wird der Wert „v1“ für die Zeile „r1“ in Tabelle „t1“ und Spalte „f1“ eingetragen.

```
hbase(main):001:0>put 't1', 'r1', 'f1:col1', 'v1'
```

## Abrufen eines Werts

Mit dem folgenden Befehl werden die Werte für Zeile „r1“ in Tabelle „t1“ abgerufen.

```
hbase(main):001:0>get 't1', 'r1'
```

## Löschen einer Tabelle

Der folgende Befehl löscht und löscht die Tabelle „t1“.

```
hbase(main):001:0>drop 'ns1:t1', false
```

Der boolesche Wert gibt an, ob Sie Ihre Tabelle archivieren möchten oder nicht. Sie können ihn also auf `true` festlegen, wenn Sie sie speichern möchten. Sie können `drop 'ns1:t1'` auch ohne booleschen Wert ausführen, um die Tabelle zu archivieren.

## Zugreifen auf HBase-Tabellen mit Hive

HBase und [Apache Hive](#) sind eng miteinander integriert, sodass Sie direkt auf in HBase gespeicherte Daten parallele Verarbeitungslasten ausführen können. Um Hive zusammen mit HBase zu nutzen, können Sie sie in der Regel auf demselben Cluster starten. Sie können Hive und HBase aber auch auf separaten Clustern starten. Wenn Sie HBase und Hive separat auf verschiedenen Clustern ausführen, kann dies die Leistung verbessern, da so jede Anwendung die Cluster-Ressourcen effizienter nutzt.

Die folgenden Verfahrensweisen zeigen, wie Sie eine Verbindung mit HBase auf einem Cluster mit Hive herstellen.

### Note

Sie können nur einen Hive-Cluster mit einem HBase-Cluster verbinden.

## So verbinden Sie Hive mit HBase

1. Erstellen Sie separate Cluster mit installiertem Hive und HBase oder erstellen Sie einen einzelnen Cluster, auf dem sowohl HBase als auch Hive installiert sind.
2. Wenn Sie separate Cluster verwenden, ändern Sie Ihre Sicherheitsgruppen so, dass HBase- und Hive-Ports zwischen diesen beiden Primärknoten geöffnet sind.
3. Verwenden Sie SSH, um eine Verbindung zum primären Knoten für den Cluster herzustellen, auf dem Hive installiert ist. Weitere Informationen finden Sie unter [Stellen Sie über SSH eine Verbindung zum Primärknoten her](#) im Verwaltungshandbuch für Amazon EMR.
4. Starten Sie die Hive-Shell mit dem folgenden Befehl.

```
hive
```

5. (Optional) Dies ist nicht erforderlich, wenn HBase und Hive sich auf demselben Cluster befinden. Verbinden Sie den HBase-Client auf Ihrem Hive-Cluster mit dem HBase-Cluster, der Ihre Daten enthält. Im folgenden Beispiel wird *Public-DNS-Name* durch den öffentlichen DNS-Namen des primären Knotens des HBase-Clusters ersetzt, zum Beispiel: `ec2-50-19-76-67.compute-1.amazonaws.com`

```
set hbase.zookeeper.quorum=public-DNS-name;
```

6. In diesem Abschnitt fahren Sie mit der Ausführung von Hive-Abfragen auf Ihre HBase-Daten fort oder sehen die nächste Vorgehensweise.

## So greifen Sie auf HBase-Daten von Hive zu

- Nach dem Herstellen der Verbindung zwischen dem Hive- und HBase-Cluster (wie in der vorherigen Vorgehensweise gezeigt) können Sie auf die Daten zugreifen, die auf dem HBase-Cluster gespeichert sind, indem Sie eine externe Tabelle in Hive erstellen.

Wenn das folgende Beispiel über die Hive-Eingabeaufforderung auf dem Primärknoten ausgeführt wird, wird eine externe Tabelle erstellt, die auf Daten verweist, die in einer HBase-Tabelle namens gespeichert sind. `inputTable` Sie können dann auf `inputTable` in Hive-Anweisungen verweisen, um Daten abzufragen und zu ändern, die im HBase-Cluster gespeichert sind.

```
set hbase.zookeeper.quorum=ec2-107-21-163-157.compute-1.amazonaws.com;

create external table inputTable (key string, value string)
  stored by 'org.apache.hadoop.hive.hbase.HBaseStorageHandler'
  with serdeproperties ("hbase.columns.mapping" = ":key,f1:col1")
  tblproperties ("hbase.table.name" = "t1");

select count(key) from inputTable ;
```

Einen fortgeschritteneren Anwendungsfall und ein Beispiel für die Kombination von HBase und Hive finden Sie im AWS Big-Data-Blogbeitrag [Kombinieren von NoSQL und massively parallel analytics using Apache HBase and Apache Hive](#) on Amazon EMR.

## Verwenden von HBase-Snapshots

HBase nutzt eine integrierte [Snapshot](#)- Funktionalität zum Erstellen kleiner Tabellensicherungen. In EMR-Clustern können diese Sicherungen mit EMRFS nach Amazon S3 exportiert werden. Sie können mithilfe der HBase-Shell einen Snapshot auf dem primären Knoten erstellen. In diesem Thema erfahren Sie, wie Sie diese Befehle interaktiv mit der Shell oder in einem Schritt ausführen, indem Sie entweder das AWS CLI oder verwenden `command-runner.jar`. AWS SDK for Java Weitere Informationen zu anderen Arten von HBase-Sicherungen finden Sie unter [HBase Backup](#) in der HBase-Dokumentation.

## Erstellen eines Snapshots mit einer Tabelle

```
hbase snapshot create -n snapshotName -t tableName
```

Verwenden von `command-runner.jar` aus dem AWS CLI:

```
aws emr add-steps --cluster-id j-2AXXXXXXGAPLF \
  --steps Name="HBase Shell Step",Jar="command-runner.jar",\
  Args=[ "hbase", "snapshot", "create", "-n", "snapshotName", "-t", "tableName" ]
```

AWS SDK for Java

```
HadoopJarStepConfig hbaseSnapshotConf = new HadoopJarStepConfig()
```

```
.withJar("command-runner.jar")  
.withArgs("hbase","snapshot","create","-n","snapshotName","-t","tableName");
```

### Note

Wenn Ihr Snapshot-Name nicht eindeutig ist, schlägt der Erstellvorgang mit dem Rückgabecode -1 oder 255 fehl. Gegebenenfalls wird aber keine Fehlermeldung mit den Details zum Fehler angezeigt. Um den gleichen Snapshot-Namen zu verwenden, löschen Sie ihn und erstellen Sie ihn dann erneut.

## Löschen eines Snapshots

```
hbase shell  
>> delete_snapshot 'snapshotName'
```

## Anzeigen von Snapshot-Informationen

```
hbase snapshot info -snapshot snapshotName
```

## Einen Snapshot nach Amazon S3 exportieren

### ⚠ Important

Wenn Sie keinen `-mappers`-Wert zum Exportieren eines Snapshots angeben, verwendet HBase eine beliebige Berechnung zur Bestimmung der Mapper-Anzahl. Dieser Wert kann je nach Tabellengröße sehr hoch sein, was sich negativ auf beim Export laufende Aufträge auswirkt. Aus diesem Grund empfehlen wir, den Parameter `-mappers`, den Parameter `-bandwidth` (der den Bandbreitenverbrauch in Megabyte pro Sekunde angibt) oder beide anzugeben, um die von der Exportoperation verwendeten Cluster-Ressourcen zu begrenzen. Alternativ können Sie die Operation zum Exportieren des Snapshots während eines Zeitraums mit geringer Auslastung ausführen.

```
hbase snapshot export -snapshot snapshotName \  
-copy-to s3://bucketName/folder -mappers 2
```

## command-runner.jar Benutzt von AWS CLI:

```
aws emr add-steps --cluster-id j-2AXXXXXXGAPLF \
--steps Name="HBase Shell Step",Jar="command-runner.jar",\
Args=[ "hbase", "snapshot", "export", "-snapshot", "snapshotName", "-copy-
to", "s3://bucketName/folder", "-mappers", "2", "-bandwidth", "50"]
```

## AWS SDK for Java:

```
HadoopJarStepConfig hbaseImportSnapshotConf = new HadoopJarStepConfig()
.withJar("command-runner.jar")
.withArgs("hbase", "snapshot", "export",
"-snapshot", "snapshotName", "-copy-to",
"s3://bucketName/folder",
"-mappers", "2", "-bandwidth", "50");
```

## Importieren eines Snapshots von Amazon S3

Obwohl es sich hier um einen Import handelt, wird auch hier die HBase-Option `export` verwendet.

```
sudo -u hbase hbase snapshot export \
-D hbase.rootdir=s3://bucketName/folder \
-snapshot snapshotName \
-copy-to hdfs://masterPublicDNSName:8020/user/hbase \
-mappers 2
```

## Verwendung command-runner.jar aus dem AWS CLI:

```
aws emr add-steps --cluster-id j-2AXXXXXXGAPLF \
--steps Name="HBase Shell Step",Jar="command-runner.jar", \
Args=["sudo", "-u", "hbase", "hbase snapshot export", "-snapshot", "snapshotName", \
"-D", "hbase.rootdir=s3://bucketName/folder", \
"-copy-to", "hdfs://masterPublicDNSName:8020/user/hbase", "-mappers", "2", "-chmod", "700"]
```

## AWS SDK for Java:

```
HadoopJarStepConfig hbaseImportSnapshotConf = new HadoopJarStepConfig()
.withJar("command-runner.jar")
.withArgs("sudo", "-u", "hbase", "hbase", "snapshot", "export", "-D", "hbase.rootdir=s3://
path/to/snapshot",
```

```
"-snapshot", "snapshotName", "-copy-to",
"hdfs://masterPublicDNSName:8020/user/hbase",
"-mappers", "2", "-chuser", "hbase");
```

## Wiederherstellen einer Tabelle von Snapshots innerhalb der HBase-Shell

```
hbase shell
>> disable tableName
>> restore_snapshot snapshotName
>> enable tableName
```

HBase unterstützt derzeit nicht alle Snapshot-Befehle in der HBase-Shell. Beispielsweise gibt es keine HBase-Befehlszeilenoption zum Wiederherstellen eines Snapshots. Deshalb müssen Sie diesen innerhalb einer Shell wiederherstellen. Dies bedeutet, dass `command-runner.jar` einen Bash-Befehl ausführen muss.

### Note

Da der hier verwendete Befehl `echo` lautet, ist es möglich, dass Ihr Shell-Befehl trotzdem fehlschlägt, selbst wenn der von Amazon EMR ausgeführte Befehl den Beendigungscode `0` zurückgibt. Überprüfen Sie die Schrittprotokolle, wenn Sie einen Shell-Befehl als Schritt ausführen möchten.

```
echo 'disable tableName; \
restore_snapshot snapshotName; \
enable tableName' | hbase shell
```

Nachfolgend finden Sie den Schritt mit der AWS CLI. Erstellen Sie zunächst die folgende `snapshot.json`-Datei:

```
[
  {
    "Name": "restore",
    "Args": ["bash", "-c", "echo '$disable \"tableName\"; restore_snapshot \
\"snapshotName\"; enable \"tableName\"; | hbase shell\""],
    "Jar": "command-runner.jar",
    "ActionOnFailure": "CONTINUE",
    "Type": "CUSTOM_JAR"
```

```
}  
]
```

```
aws emr add-steps --cluster-id j-2AXXXXXXGAPLF \  
--steps file:///./snapshot.json
```

AWS SDK for Java:

```
HadoopJarStepConfig hbaseRestoreSnapshotConf = new HadoopJarStepConfig()  
.withJar("command-runner.jar")  
.withArgs("bash", "-c", "echo '$disable \"tableName\"; restore_snapshot \"snapshotName  
\"; enable \"snapshotName\"' | hbase shell");
```

## Konfigurieren von HBase

Obwohl die Standardeinstellungen von HBase für die meisten Anwendungen ausreichen sollten, können Sie Ihre HBase-Konfigurationseinstellungen ändern. Verwenden Sie hierfür Eigenschaften der HBase-Konfigurationsklassifizierungen. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Im folgenden Beispiel wird ein Cluster mit einem alternativen HBase-Stammverzeichnis basierend auf der Konfigurationsdatei `myConfig.json` erstellt, die in Amazon S3 gespeichert ist.

### Note

Linux-Zeilenumbruchzeichen (`\`) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (`^`).

```
aws emr create-cluster --release-label emr-7.1.0 --applications Name=HBase \  
--instance-type m5.xlarge --instance-count 3 --configurations https://s3.amazonaws.com/  
mybucket/myfolder/myConfig.json
```

Die Datei `myConfig.json` gibt die `hbase.rootdir`-Eigenschaft für die `hbase-site`-Konfigurationsklassifizierung an, wie im folgenden Beispiel gezeigt. Ersetzen Sie *IP-xxx-xx-xx-xxx.ec2.internal* durch den internen DNS-Hostnamen des primären Knotens des Clusters.



```
[
  {
    "Classification": "hbase-site",
    "Properties": {
      "hbase.rootdir": "hdfs://ip-XXX-XX-XX-XXX.ec2.internal:8020/user/
myCustomHBaseDir"
    }
  }
]
```

### Note

Ab Amazon-EMR-Version 5.21.0 können Sie Cluster-Konfigurationen überschreiben und zusätzliche Konfigurationsklassifikationen für jede Instance-Gruppe in einem ausgeführten Cluster angeben. Dazu verwenden Sie die Amazon EMR-Konsole, das AWS Command Line Interface (AWS CLI) oder das AWS SDK. Weitere Informationen finden Sie unter [Angeben einer Konfiguration für eine Instance-Gruppe in einem aktiven Cluster](#).

## Änderungen an der Arbeitsspeicherzuordnung in YARN

HBase wird nicht als YARN Anwendung ausgeführt. Folglich ist es erforderlich, den Arbeitsspeicher, der YARN und seinen Anwendungen zugewiesen ist, neu zu berechnen, was zu einer Reduzierung des Gesamtspeichers führt, der YARN zur Verfügung steht, wenn HBase installiert ist. Sie sollten dies berücksichtigen, wenn Sie planen, YARN-Anwendungen und HBase auf dem gleichen Cluster anzusiedeln. Für Instance-Typen mit weniger als 64 GB Arbeitsspeicher steht die Hälfte des Speichers zur Verfügung NodeManager, der dann der RegionServer HBase zugewiesen wird. Bei Instance-Typen mit mehr als 64 GB Arbeitsspeicher ist der RegionServer HBase-Speicher auf 32 GB begrenzt. Als allgemeine Regel gilt, dass der YARN-Einstellungsspeicher ein Vielfaches des MapReduce Reducer-Task-Speichers ist.

Die Tabellen unter [Standardwerte für die Aufgabenkonfigurationseinstellungen](#) zeigen die Änderungen an den YARN-Einstellungen basierend auf dem Arbeitsspeicher, der für HBase erforderlich ist.

## HBase-Port-Nummern

Einige Port-Nummern, die für HBase gewählt werden, unterscheiden sich vom Standard. Die folgenden Schnittstellen und Ports sind für HBase in Amazon EMR verfügbar.

## HBase-Ports

Schnittstelle	Port	Protokoll
HMaster	16000	TCP
HMaster-UI	16010	HTTP
RegionServer	16020	TCP
RegionServer Informationen	16030	HTTP
REST-Server	8070	HTTP
REST-UI	8085	HTTP
Thrift-Server	9090	TCP
Thrift-Server-UI	9095	HTTP

### Important

In Amazon-EMR-Version 4.6.0 und höher ist der `kms-http-port` 9700 und der `kms-admin-port` 9701.

## Zu optimierende HBase-Standorteinstellungen

Sie können einzelne oder alle HBase-Standorteinstellungen konfigurieren, um den HBase-Cluster für den Workload Ihrer Anwendung zu optimieren. Wir empfehlen die folgenden Einstellungen als Ausgangspunkt.

### `zookeeper.session.timeout`

Der Timeout-Standardwert beträgt 40 Sekunden (40 000 ms). Wenn ein Regionsserver ausfällt, gibt dieser Wert an, wie lange es dauert, bis der Master-Server die Abwesenheit des Regionsservers erkennt und mit der Wiederherstellung beginnt. Um die Wiederherstellung des Master-Servers zu beschleunigen, können Sie diesen Wert auf einen kürzeren Zeitraum einstellen. Im folgenden Beispiel werden 30 Sekunden oder 30 000 ms verwendet:

```
[
  {
    "Classification":"hbase-site",
    "Properties": {
      "zookeeper.session.timeout": "30000"
    }
  }
]
```

## hbase.regionserver.handler.count

Dadurch wird die Anzahl der Threads definiert, die der Regionsserver offen hält, um Tabellenanforderungen zu verarbeiten. Die Standardeinstellung 10 ist niedrig, damit Benutzer ihre Regionsserver nicht überlasten, wenn sie zu große Schreibpuffer mit vielen gleichzeitigen Clients verwenden. In der Regel sollte diese Zahl möglichst gering gehalten werden, wenn die Nutzdaten pro Anforderung den MB-Bereich erreicht (große Puts, Scans mit großem Zwischenspeicher), bzw. hoch, wenn die Nutzdatenlast gering ist (Gets, kleine Puts, ICVs, Löschvorgänge). Im folgenden Beispiel wird die Anzahl der offenen Threads auf 30 erhöht:

```
[
  {
    "Classification":"hbase-site",
    "Properties": {
      "hbase.regionserver.handler.count": "30"
    }
  }
]
```

## hbase.hregion.max.filesize

Dieser Parameter steuert die Größe (in Byte) der einzelnen Regionen. Standardmäßig ist der Wert eingestellt 1073741824. Wenn Sie große Datenmengen in Ihren HBase-Cluster schreiben und dies häufige Aufteilungen verursacht, können Sie diesen Wert erhöhen, um einzelne Regionen zu vergrößern. Dadurch werden zwar Aufteilungen reduziert, es braucht jedoch mehr Zeit, um die Workloads in den Regionen zwischen den Servern zu verteilen.

```
[
  {
    "Classification":"hbase-site",
```

```
"Properties": {  
  "hbase.hregion.max.filesize": "1073741824"  
}  
}  
]
```

## hbase.hregion.memstore.flush.size

Dieser Parameter steuert die maximale Größe des Memstore (in Byte), bevor er auf den Datenträger ausgelagert wird. Standardmäßig ist dieser 134217728. Wenn Ihr Workload aus Schreibvorgängen in kurzen Spitzen besteht, sollten Sie dieses Limit erhöhen, damit alle Schreibvorgänge während der Spitzenauslastung im Arbeitsspeicher verbleiben und zu einem späteren Zeitpunkt auf den Datenträger ausgelagert werden. Dies kann die Leistung bei Auslastungsspitzen verbessern.

```
[  
  {  
    "Classification": "hbase-site",  
    "Properties": {  
      "hbase.hregion.memstore.flush.size": "134217728"  
    }  
  }  
]
```

## Anzeigen der HBase-Benutzeroberfläche

### Note

Die HBase-Benutzeroberfläche verwendet standardmäßig unsichere HTTP-Verbindungen. Um sicheres HTTP (HTTPS) zu aktivieren, setzen Sie die `hbase.ssl.enabled`-Eigenschaft für die `hbase-site`-Klassifizierung in Ihrer [HBase-Konfiguration](#) auf `true`. Weitere Informationen zur Verwendung von sicherem HTTP (HTTPS) für die HBase-Weboberfläche finden Sie im [Apache-HBase-Referenzhandbuch](#).

HBase bietet eine webbasierte Benutzeroberfläche, über die Sie Ihren HBase-Cluster überwachen können. Wenn Sie HBase auf Amazon EMR ausführen, wird das Webinterface auf dem primären Knoten ausgeführt und kann mithilfe von Portweiterleitung, auch bekannt als Erstellen eines SSH-Tunnels, angezeigt werden.

So zeigen Sie die HBase-Benutzeroberfläche an

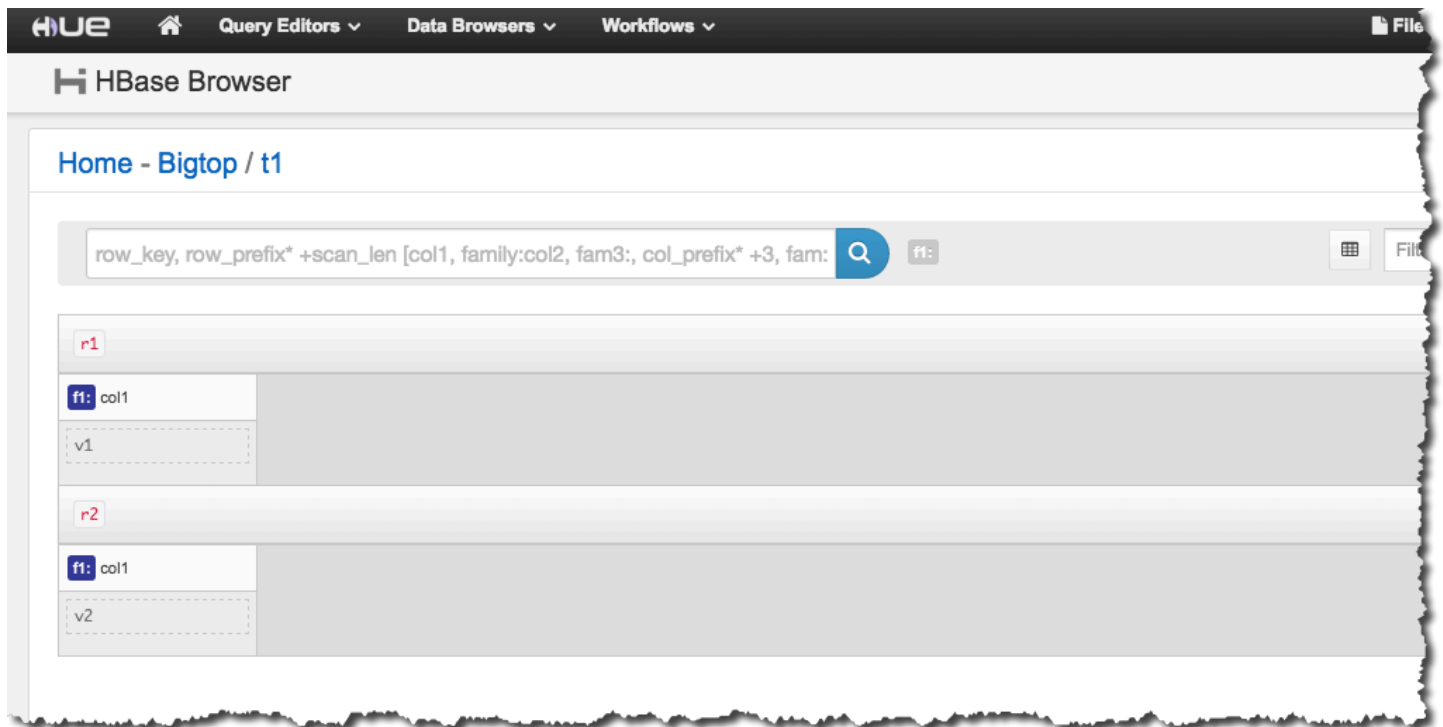
1. Verwenden Sie SSH, um einen Tunnel zum Primärknoten zu tunneln und eine sichere Verbindung herzustellen. Weitere Informationen finden Sie unter [Option 2, Teil 1: Einen SSH-Tunnel zum Primärknoten mithilfe dynamischer Portweiterleitung einrichten](#) im Amazon EMR Management Guide.
2. Installieren Sie einen Webbrowser mit einem Proxy-Tool, z. B. dem FoxyProxy Plug-in für Firefox, um einen SOCKS-Proxy für AWS Domains zu erstellen. Weitere Informationen finden Sie unter [Option 2, Teil 2: Proxyeinstellungen konfigurieren, um Websites anzuzeigen, die auf dem primären Knoten gehostet werden](#), im Amazon EMR Management Guide.
3. Wenn der Proxy eingerichtet und die SSH-Verbindung geöffnet ist, können Sie die HBase-Benutzeroberfläche aufrufen, indem Sie ein Browserfenster mit `http://:16010/master-status` öffnen *master-public-dns-name*, in dem sich die öffentliche DNS-Adresse des Primärknotens *master-public-dns-name* des Clusters befindet.

The screenshot shows the Apache HBase web interface. At the top, there is a navigation bar with the following links: Home, Table Details, Local Logs, Log Level, Debug Dump, Metrics Dump, and HBase Configuration. Below the navigation bar, the page title is "Master" followed by a blurred IP address and ".ec2.internal".

Under the "Master" heading, there is a section for "Region Servers". This section has several tabs: "Base Stats" (selected), "Memory", "Requests", "Storefiles", and "Compactions". Below the tabs is a table with the following columns: "ServerName", "Start time", "Version", "Requests Per Second", and "Num. P".

ServerName	Start time	Version	Requests Per Second	Num. P
.ec2.internal,16020,1461165084992	Wed Apr 20 15:11:24 UTC 2016	1.2.0	0	1
.ec2.internal,16020,1461165087881	Wed Apr 20 15:11:27 UTC 2016	1.2.0	0	2
Total:2			0	3

Sie können HBase auch in Hue anzeigen. Das folgende Beispiel zeigt die Tabelle `t1`, erstellt in [Verwenden der HBase-Shell](#):



Weitere Informationen zu Hue finden Sie unter [Hue](#).

## Anzeigen der HBase-Protokolldateien

Im Rahmen seines Betriebs schreibt HBase Protokolldateien mit Informationen zu Konfigurationseinstellungen, Daemon-Aktionen und Ausnahmen. Diese Protokolldateien können zum Debuggen von Problemen mit HBase sowie für die Verfolgung der Leistung hilfreich sein.

Wenn Sie Ihren Cluster so konfigurieren, dass die Protokolldateien dauerhaft in Amazon S3 gespeichert werden, sollten Sie wissen, dass Protokolle alle fünf Minuten in Amazon S3 geschrieben werden. Deshalb kann es zu einer kleinen Verzögerung kommen, bis die neuesten Protokolldateien verfügbar sind.

Um HBase-Protokolle auf dem primären Knoten anzuzeigen

- Sie können die aktuellen HBase-Protokolle anzeigen, indem Sie SSH verwenden, um eine Verbindung zum primären Knoten herzustellen, und zum Verzeichnis navigieren. `/var/log/hbase` Diese Protokolle sind nicht verfügbar, nachdem der Cluster beendet wurde, es sei denn, Sie aktivieren beim Starten des Clusters die Protokollierung nach Amazon S3.

## So zeigen Sie HBase-Protokolle in Amazon S3 an

- Um auf HBase-Protokolle und andere Cluster-Protokolle in Amazon S3 zuzugreifen und sie auch noch verfügbar zu haben, nachdem der Cluster beendet wurde, müssen Sie beim Erstellen des Clusters einen Amazon-S3-Bucket angeben, um diese Protokolle zu empfangen. Dies erfolgt mithilfe der Option `--log-uri`. Weitere Informationen zur Aktivierung der Protokollierung für Ihren Cluster finden Sie unter [Konfiguration von Protokollierung und Debugging \(Optional\)](#) im Verwaltungshandbuch für Amazon EMR.

## Überwachen von HBase mit Ganglia

Das Open-Source-Projekt Ganglia ist ein skalierbares, verteiltes System zur Überwachung von Clustern und Grids, das zugleich die Auswirkungen auf die Leistung minimiert. Wenn Sie Ganglia in Ihrem Cluster aktivieren, können Sie Berichte erstellen und die Leistung des Clusters als Ganzes betrachten. Ebenso können Sie die Leistung einzelner Knoten-Instances überprüfen. Weitere Informationen zum Open-Source-Projekt Ganglia finden Sie unter <http://ganglia.info/>. Weitere Informationen zur Verwendung von Ganglia mit Amazon-EMR-Clustern finden Sie unter [Ganglia](#).

Nachdem der Cluster mit konfigurierter Ganglia gestartet wurde, können Sie über die grafische Oberfläche, die auf dem primären Knoten ausgeführt wird, auf die Ganglia-Diagramme und -Berichte zugreifen.

Ganglia speichert Protokolldateien auf dem primären Knoten im Verzeichnis `/mnt/var/lib/ganglia/rrds/`. Frühere Versionen von Amazon EMR speichern Protokolldateien möglicherweise im `/var/log/ganglia/rrds/`-Verzeichnis.

Um einen Cluster für Ganglia und HBase zu konfigurieren, verwenden Sie AWS CLI

- Verwenden Sie einen `create-cluster`-Befehl ähnlich dem folgenden:

```
aws emr create-cluster --name "Test cluster" --release-label emr-7.1.0 \  
--applications Name=HBase Name=Ganglia --use-default-roles \  
--ec2-attributes KeyName=myKey --instance-type m5.xlarge \  
--instance-count 3
```

**Note**

Wenn die Standard-Amazon-EMR-Servicerolle und das Amazon-EC2-Instance-Profil nicht vorhanden sind, tritt ein Fehler auf. Verwenden Sie den Befehl `aws emr create-default-roles`, um die Rolle und das Profil zu erstellen, und versuchen Sie es erneut.

Weitere Informationen finden Sie unter [Amazon-EMR-Befehle in der AWS CLI](#).

So zeigen Sie HBase-Metriken auf der Weboberfläche von Ganglia an

1. Verwenden Sie SSH, um einen Tunnel in den Primärknoten zu tunneln und eine sichere Verbindung herzustellen. Weitere Informationen finden Sie unter [Option 2, Teil 1: Einen SSH-Tunnel zum Primärknoten mithilfe dynamischer Portweiterleitung einrichten](#) im Amazon EMR Management Guide.
2. Installieren Sie einen Webbrowser mit einem Proxy-Tool, z. B. dem FoxyProxy Plug-in für Firefox, um einen SOCKS-Proxy für AWS Domains zu erstellen. Weitere Informationen finden Sie unter [Option 2, Teil 2: Proxyeinstellungen konfigurieren, um Websites anzuzeigen, die auf dem primären Knoten gehostet](#) werden, im Amazon EMR Management Guide.
3. Wenn der Proxy eingerichtet und die SSH-Verbindung geöffnet ist, können Sie die Ganglia-Metriken anzeigen, indem Sie ein Browserfenster mit `http://master-public-dns-name/ganglia/` öffnen, in dem sich die öffentliche DNS-Adresse des Masterservers im HBase-Cluster *master-public-dns-name* befindet.

Um Ganglia-Protokolldateien auf dem primären Knoten anzuzeigen

- Wenn der Cluster noch läuft, können Sie auf die Protokolldateien zugreifen, indem Sie SSH verwenden, um eine Verbindung zum primären Knoten herzustellen und zum Verzeichnis zu navigieren. `/mnt/var/lib/ganglia/rrds/` Navigieren Sie für EMR 3.x zum Verzeichnis `/var/log/ganglia/rrds`. Weitere Informationen finden Sie unter [Stellen Sie über SSH eine Verbindung zum Primärknoten her](#) im Verwaltungshandbuch für Amazon EMR.



So zeigen Sie Ganglia-Protokolldateien in Amazon S3 an

- Die Ganglia-Protokolldateien werden nicht automatisch nach Amazon S3 geschrieben, auch wenn Sie die Protokollierung für den Cluster aktivieren. Um die Ganglia-Protokolldateien in Amazon S3 anzuzeigen, müssen Sie die Protokolle manuell per Push von `/mnt/var/lib/ganglia/rrds/` in den S3-Bucket übertragen.

## Migrieren von vorherigen HBase-Versionen

Informationen zum Migrieren von Daten aus einer vorherigen HBase-Version finden Sie unter [Upgrading](#) und [HBase version number and compatibility](#) im Apache HBase-Referenzleitfaden. Möglicherweise müssen Sie besonders auf die Anforderungen in Bezug auf Upgrades von HBase-Versionen vor Version 1.0 achten.

## Versionsverlauf von HBase

In der folgenden Tabelle sind die Version von HBase aufgeführt, die in jeder Release-Version von Amazon EMR enthalten ist, zusammen mit den Komponenten, die mit der Anwendung installiert wurden. Informationen zu den Komponentenversionen in den einzelnen Versionen finden Sie im Abschnitt Komponentenversion für Ihre Version in [Amazon-EMR-7.x-Versionen](#), [Amazon-EMR-6.x-Versionen](#) oder [Amazon-EMR-5.x-Versionen](#).

Die Versionsinformationen HBase

Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-5.36.2	1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hma

Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
		ster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-7.1.0	2.4.17	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-wal-cli, hadoop-client, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-7.0.0	2.4.17	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-wal-cli, hadoop-client, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-6.15.0	2.4.17	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-wal-cli, hadoop-client, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-ma-pred, hadoop-yarn-nodema-nager, hadoop-yarn-resour-cemanager, hadoop-yarn-timeline-server, hbase-hma-ster, hbase-client, hbase-reg-ion-server, hbase-rest-server, hbase-thrift-server, hbase-ope-rator-tools, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-6.14.0	2.4.17	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-wal-cli, hadoop-client, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop- kms-server, hadoop-ma pred, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hbase-hma ster, hbase-client, hbase-reg ion-server, hbase-rest-server, hbase-thrift-server, hbase-ope rator-tools, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-6.13.0	2.4.17	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-wal-cli, hadoop-client, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-ma-pred, hadoop-yarn-nodema-nager, hadoop-yarn-resour-cemanager, hadoop-yarn-timeline-server, hbase-hma-ster, hbase-client, hbase-reg-ion-server, hbase-rest-server, hbase-thrift-server, hbase-ope-rator-tools, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-6.12.0	2.4.17	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-wal-cli, hadoop-client, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-6.11.1	2.4,15	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-wal-cli, hadoop-client, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-ma-pred, hadoop-yarn-nodema-nager, hadoop-yarn-resour-cemanager, hadoop-yarn-timeline-server, hbase-hma-ster, hbase-client, hbase-reg-ion-server, hbase-rest-server, hbase-thrift-server, hbase-ope-rator-tools, zookeeper-client, zookeeper-server



Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-6.11.0	2.4,15	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-wal-cli, hadoop-client, hadoop-hdfs-datano- de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop- httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodema- nager, hadoop-yarn-resourceman- ager, hadoop-yarn-timeline-server, hbase-hma- ster, hbase-client, hbase-reg- ion-server, hbase-rest-server, hbase-thrift-server, hbase-ope- rator-tools, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-6.10.1	2.4,15	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-wal-cli, hadoop-client, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-ma-pred, hadoop-yarn-nodema-nager, hadoop-yarn-resour-cemanager, hadoop-yarn-timeline-server, hbase-hma-ster, hbase-client, hbase-reg-ion-server, hbase-rest-server, hbase-thrift-server, hbase-ope-rator-tools, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-6.10.0	2.4,15	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-wal-cli, hadoop-client, hadoop-hdfs-datano- de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop- httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourceman- ager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-6.9.1	2.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server
emr-6.9.0	2.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-6.8.1	2.4.12	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server
emr-6.8.0	2.4.12	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-6.7.0	2.4.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server
emr-5.36.1	1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-5.36.0	1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-6.6.0	2.4.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-5.35.0	1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-6.5.0	2.4.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server



Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-6.4.0	2.4.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-6.3.1	2.2.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-6.3.0	2.2.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-6.2.1	2.2.6-amzn-0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-6.2.0	2.2.6-amzn-0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-6.1.1	2.2.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-6.1.0	2.2.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-6.0.1	2.2.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-6.0.0	2.2.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.34.0	1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-5.33.1	1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.33.0	1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-5.32.1	1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.32.0	1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-5.31.1	1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.31.0	1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server



Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-5.30.2	1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.30.1	1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-5.30.0	1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.29.0	1.4.10	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-5.28.1	1.4.10	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.28.0	1.4.10	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-5.27.1	1.4.10	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.27.0	1.4.10	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-5.26.0	1.4.10	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.25.0	1.4.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-5.24.1	1.4.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.24.0	1.4.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-5.23.1	1.4.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.23.0	1.4.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-5.22.0	1.4.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.21.2	1.4.8	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server



Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-5.21.1	1.4.8	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.21.0	1.4.8	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-5.20.1	1.4.8	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.20.0	1.4.8	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-5.19.1	1.4.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.19.0	1.4.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-5.18.1	1.4.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.18.0	1.4.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-5.17.2	1.4.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.17.1	1.4.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-5.17.0	1.4.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.16.1	1.4.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-5.16.0	1.4.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.15.1	1.4.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-5.15.0	1.4.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.14.2	1.4.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server



Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-5.14.1	1.4.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.14.0	1.4.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-5.13.1	1.4.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.13.0	1.4.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-5.12.3	1.4.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.12.2	1.4.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-5.12.1	1.4.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.12.0	1.4.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-5.11.4	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.11.3	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-5.11.2	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.11.1	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-5.11.0	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.10.1	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-5.10.0	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.9.1	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server



Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-5.9.0	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.8.3	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-5.8.2	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.8.1	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-5.8.0	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.7.1	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-5.7.0	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.6.1	1.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-5.6.0	1.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.5.4	1.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-5.5.3	1.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.5.2	1.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-5.5.1	1.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.5.0	1.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-5.4.1	1.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.4.0	1.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server



Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-5.3.2	1.2.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.3.1	1.2.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-5.3.0	1.2.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.2.3	1.2.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-5.2.2	1.2.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.2.1	1.2.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-5.2.0	1.2.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.1.1	1.2.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-5.1.0	1.2.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.0.3	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-5.0.2	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.0.1	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-5.0.0	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-4.9.6	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-4.9.5	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-4.9.4	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server



Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-4.9.3	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-4.9.2	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-4.9.1	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-4.8.5	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-4.8.4	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-4.8.3	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-4.8.2	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-4.8.1	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-4.8.0	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-4.7.4	1.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-4.7.3	1.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-4.7.2	1.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-4.7.1	1.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-4.7.0	1.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	HBase-Version	Mit HBase installierte Komponenten
emr-4.6.1	1.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-4.6.0	1.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server



# Apache HCatalog

HCatalog ist ein Tool, mit dem Sie in Pig-, Spark SQL- und/oder benutzerdefinierten Anwendungen auf Hive-Metastore-Tabellen zugreifen können. MapReduce HCatalog verfügt über eine REST-Schnittstelle und einen Befehlszeilen-Client, mit dem Sie Tabellen erstellen oder andere Operationen ausführen können. Anschließend schreiben Sie Ihre Anwendungen, um auf die Tabellen mit HCatalog-Bibliotheken zuzugreifen. Weitere Informationen finden Sie unter [Verwenden von HCatalog](#). HCatalog ist in Amazon-EMR-Version 4.4.0 und höher enthalten.

HCatalog auf Amazon EMR, Version 5.8.0 und höher, unterstützt die Verwendung von AWS Glue Data Catalog als Metastore für Hive. Weitere Informationen finden Sie unter [Verwenden des AWS Glue-Datenkatalogs als Metastore für Hive](#).

Die folgende Tabelle listet die Version von HCatalog auf, die in der neuesten Version der Amazon-EMR-6.x-Serie enthalten ist, zusammen mit den Komponenten, die Amazon EMR mit HCatalog installiert.

Die Version der Komponenten, die mit HCatalog in dieser Version installiert wurden, finden Sie unter [Komponentenversionen der Version 6.15.0](#).

HCatalog-Versionsinformationen für emr-6.15.0

Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-6.15.0	HCatalog 3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Die folgende Tabelle listet die Version von HCatalog auf, die in der neuesten Version der Amazon-EMR-5.x-Serie enthalten ist, zusammen mit den Komponenten, die Amazon EMR mit HCatalog installiert.

Informationen zur Version der Komponenten, die in dieser Version mit HCatalog installiert wurden, finden Sie unter [Komponentenversionen von Version 5.36.2](#).

#### HCatalog-Versionsinformationen für emr-5.36.2

Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-5.36.2	HCatalog 2.3.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

#### Themen

- [Erstellen eines Clusters mit HCatalog](#)
- [Verwenden von HCatalog](#)
- [Beispiel: Erstellen einer HCatalog-Tabelle und Ausführen eines Schreibvorgangs mit Pig](#)
- [Versionsverlauf von HCatalog](#)

## Erstellen eines Clusters mit HCatalog

Obwohl HCatalog im Hive-Projekt enthalten ist, müssen Sie es als eigene Anwendung installieren.

## So starten Sie einen Cluster mit HCatalog über die Konsole

Im folgenden Verfahren wird ein Cluster mit installiertem HCatalog erstellt. Weitere Informationen zum Erstellen von Clustern mit der Konsole, einschließlich Erweiterte Optionen, finden Sie unter [Cluster planen und konfigurieren](#) im Verwaltungshandbuch für Amazon EMR.

1. Öffnen Sie die Amazon EMR-Konsole unter <https://console.aws.amazon.com/emr>.
2. Wählen Sie Create cluster (Cluster erstellen) aus, um Quick Create (Schnellerstellung) zu verwenden.
3. Wählen Sie im Feld Software Configuration (Softwarekonfiguration) die Option Amazon Release Version emr-4.4.0 oder höher aus.
4. Wählen Sie im Feld Select Applications (Anwendungen auswählen) entweder All Applications (Alle Anwendungen) oder HCatalog aus.
5. Wählen Sie nach Bedarf weitere Optionen und anschließend Create cluster (Cluster erstellen) aus.

Um einen Cluster mit HCatalog zu starten, verwenden Sie AWS CLI

- Erstellen Sie den Cluster mit dem folgenden Befehl:

### Note

Linux-Zeilenfortsetzungszeichen (\) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (^).

```
aws emr create-cluster --name "Cluster with Hcat" --release-label emr-7.1.0 \  
--applications Name=HCatalog --ec2-attributes KeyName=myKey \  
--instance-type m5.xlarge --instance-count 3 --use-default-roles
```

## Verwenden von HCatalog

Sie können HCatalog in verschiedenen Anwendungen verwenden, die den Hive-Metastore nutzen. Die Beispiele in diesem Abschnitt zeigen, wie Sie eine Tabelle erstellen und sie im Kontext von Pig und Spark SQL verwenden.

## Deaktivieren der Direct-Write-Funktion bei der Verwendung von HCatalog HStorer

Wenn eine Anwendung [H](#) verwendet CatStorer, um in eine in Amazon S3 gespeicherte HCatalog-Tabelle zu schreiben, deaktivieren Sie die Direktschreibfunktion von Amazon EMR. Deaktivieren Sie Direct-Write beispielsweise bei Verwendung des Pig STORE-Befehls oder bei Ausführung von Sqoop-Aufträgen, bei denen HCatalog-Tabellen in Amazon S3 geschrieben werden. Sie können die Direct Write-Funktion deaktivieren, indem Sie die Konfigurationen `mapred.output.direct.NativeS3FileSystem` und `mapred.output.direct.EmrFileSystem` auf `false` setzen. Das folgende Beispiel zeigt, die diese Konfigurationen mit Java festgelegt werden.

```
Configuration conf = new Configuration();
conf.set("mapred.output.direct.NativeS3FileSystem", "false");
conf.set("mapred.output.direct.EmrFileSystem", "false");
```

## Erstellen einer Tabelle mit der HCat CLI und Verwenden der entsprechenden Daten in Pig

Erstellen Sie das folgende Skript, `impressions.q`, auf Ihrem Cluster:

```
CREATE EXTERNAL TABLE impressions (
  requestBeginTime string, adId string, impressionId string, referrer string,
  userAgent string, userCookie string, ip string
)
PARTITIONED BY (dt string)
ROW FORMAT
  serde 'org.apache.hive.hcatalog.data.JsonSerDe'
  with serdeproperties ( 'paths'='requestBeginTime, adId, impressionId, referrer,
userAgent, userCookie, ip' )
LOCATION 's3://[your region].elasticmapreduce/samples/hive-ads/tables/impressions/';
ALTER TABLE impressions ADD PARTITION (dt='2009-04-13-08-05');
```

Führen Sie das Skript mit der HCat CLI aus:

```
% hcat -f impressions.q
Logging initialized using configuration in file:/etc/hive/conf.dist/hive-
log4j.properties
OK
Time taken: 4.001 seconds
```

```
OK
Time taken: 0.519 seconds
```

Öffnen Sie die Grunt-Shell und greifen Sie auf die Daten in impressions zu:

```
% pig -useHCatalog -e "A = LOAD 'impressions' USING
  org.apache.hive.hcatalog.pig.HCatLoader();
B = LIMIT A 5;
dump B;"
<snip>
(1239610346000,m9nwd067Nx6q2kI25qt50n7peICfUM,omkxkaRpNhGPDucAiBErSh1cs0MThC,cartoonnetwork.com
  (compatible; MSIE 7.0; Windows NT 6.0; FunWebProducts; GTB6; SLCC1; .NET CLR
  2.0.50727; Media Center PC
  5.0; .NET,wcVWWTascoPbGt6bdqDbuWTPPHgOPs,69.191.224.234,2009-04-13-08-05)
(1239611000000,NjriQjd0DgWBKnkGJUP6GNTbDeK4An,AWtXPkfaWGOaNeL900sFU8Hcj6eLHT,cartoonnetwork.com
  (compatible; MSIE 7.0; Windows NT 5.1; GTB6; .NET CLR
  1.1.4322),0aMU1F2gE4CtADVHAbKjRRks5kIgg,57.34.133.110,2009-04-13-08-05)
(1239610462000,Irpv3oiu0I5QNQiWSSITishrLdo9cM1,i1LDq44LRSJF0hbmhB8Gk7k9gMWtBq,cartoonnetwork.com
  (compatible; MSIE 6.0; Windows NT 5.2; SV1; .NET CLR 1.1.4322;
  InfoPath.1),QsB3wkLR4JAIut4Uq6FNFQIR1rCVwU,42.174.193.253,2009-04-13-08-05)
(1239611007000,q2Awfnpe0JAvhInaIp0VGx9Kts0oPO,s3HvTf1PB8JIE0IuM6h0EebWwP0tJV,cartoonnetwork.com
  (compatible; MSIE 6.0; Windows NT 5.2; SV1; .NET CLR 1.1.4322;
  InfoPath.1),QsB3wkLR4JAIut4Uq6FNFQIR1rCVwU,42.174.193.253,2009-04-13-08-05)
(1239610398000,c362vpAB0soPKGHRs43cj6TRwNe0Gn,jeas5nXbQInGAgFB8jlkhnprN6cMw7,cartoonnetwork.com
  (compatible; MSIE 8.0; Windows NT 5.1; Trident/4.0; GTB6; .NET CLR
  1.1.4322),k96n5PnUmwHKfiUI0TFP0TNMfADgh9,51.131.29.87,2009-04-13-08-05)
7120 [main] INFO  org.apache.pig.Main - Pig script completed in 7 seconds and 199
  milliseconds (7199 ms)
16/03/08 23:17:10 INFO pig.Main: Pig script completed in 7 seconds and 199 milliseconds
  (7199 ms)
```

## Zugreifen auf die Tabelle mit Spark SQL

Dieses Beispiel erstellt einen Spark DataFrame aus der Tabelle, die im ersten Beispiel erstellt wurde, und zeigt die ersten 20 Zeilen:

```
% spark-shell --jars /usr/lib/hive-hcatalog/share/hcatalog/hive-hcatalog-core-1.0.0-
  amzn-3.jar
<snip>
scala> val hiveContext = new org.apache.spark.sql.hive.HiveContext(sc);
scala> val df = hiveContext.sql("SELECT * FROM impressions")
scala> df.show()
```

&lt;snip&gt;

16/03/09 17:18:46 INFO DAGScheduler: ResultStage 0 (show at <console>:32) finished in 10.702 s

16/03/09 17:18:46 INFO DAGScheduler: Job 0 finished: show at <console>:32, took 10.839905 s

```

+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+
|requestbegintime|          adid|      impressionid|      referrer|
|  useragent|      usercookie|      ip|      dt|
+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+
|  1239610346000|m9nwdo67Nx6q2kI25...|omkxkaRpNhGPDucAi...|cartoonnetwork.com|
Mozilla/4.0 (comp...|wcVWWTascoPbGt6bd...|69.191.224.234|2009-04-13-08-05|
|  1239611000000|NjriQjd0DgWBKnkGJ...|AWtXPkfaWG0aNeL90...|cartoonnetwork.com|
Mozilla/4.0 (comp...|0aMU1F2gE4CtADVHA...| 57.34.133.110|2009-04-13-08-05|
|  1239610462000|Irpv3oiu0I5QNQiW...|i1LDq44LRSJF0hbmh...|cartoonnetwork.com|
Mozilla/4.0 (comp...|QSb3wkLR4JAIut4Uq...|42.174.193.253|2009-04-13-08-05|
|  1239611007000|q2Awfnp0JAVhInaI...|s3HvTfLPB8JIE0IuM...|cartoonnetwork.com|
Mozilla/4.0 (comp...|QSb3wkLR4JAIut4Uq...|42.174.193.253|2009-04-13-08-05|
|  1239610398000|c362vpAB0soPKGHS...|jeas5nXbQInGAgFB8...|cartoonnetwork.com|
Mozilla/4.0 (comp...|k96n5PnUmwHKfiUI0...| 51.131.29.87|2009-04-13-08-05|
|  1239610600000|cjBTpruoaiEtqLuMX...|XwlohBSs8Ipxs1bRa...|cartoonnetwork.com|
Mozilla/4.0 (comp...|k96n5PnUmwHKfiUI0...| 51.131.29.87|2009-04-13-08-05|
|  1239610804000|Ms3eJHNAEItpxvimd...|4SIj4pGmgVLL625BD...|cartoonnetwork.com|
Mozilla/4.0 (comp...|k96n5PnUmwHKfiUI0...| 51.131.29.87|2009-04-13-08-05|
|  1239610872000|h5bccHX6wJReDi1jL...|EFAWiiBdVfnxwAMWP...|cartoonnetwork.com|
Mozilla/4.0 (comp...|k96n5PnUmwHKfiUI0...| 51.131.29.87|2009-04-13-08-05|
|  1239610365000|874NBpGmxNFfxEPKM...|xSvE4XtGbdtXPF2Lb...|cartoonnetwork.com|
Mozilla/5.0 (Maci...|eWDEVVUphlnRa273j...| 22.91.173.232|2009-04-13-08-05|
|  1239610348000|X8gISpUTSqh1A5reS...|TrFblGT99AgE75vuj...|      corriere.it|
Mozilla/4.0 (comp...|tX1sMpnhJUhmAF7AS...| 55.35.44.79|2009-04-13-08-05|
|  1239610743000|kbKreLWB6QVueFrDm...|kVnxx9Ie2i30LTxFj...|      corriere.it|
Mozilla/4.0 (comp...|tX1sMpnhJUhmAF7AS...| 55.35.44.79|2009-04-13-08-05|
|  1239610812000|9lx0SRpEi3bmEeTCu...|1B2sff99AEIwSuLVV...|      corriere.it|
Mozilla/4.0 (comp...|tX1sMpnhJUhmAF7AS...| 55.35.44.79|2009-04-13-08-05|
|  1239610876000|lijjmCf2kuxfBTnjL...|AjvufgUtakUFcsIM9...|      corriere.it|
Mozilla/4.0 (comp...|tX1sMpnhJUhmAF7AS...| 55.35.44.79|2009-04-13-08-05|
|  1239610941000|t8t8trgjNRPIlmxuD...|agu2u2TCdqWP08rAA...|      corriere.it|
Mozilla/4.0 (comp...|tX1sMpnhJUhmAF7AS...| 55.35.44.79|2009-04-13-08-05|
|  1239610490000|OGRLPVNGxiGgrCmWL...|mJg2raBUpPrC80lUm...|      corriere.it|
Mozilla/4.0 (comp...|r2k96t1CNjSU9fJKN...| 71.124.66.3|2009-04-13-08-05|
|  1239610556000|OnJID12x0RXKPUgrD...|P7Pm2mPdW6w08KA3R...|      corriere.it|
Mozilla/4.0 (comp...|r2k96t1CNjSU9fJKN...| 71.124.66.3|2009-04-13-08-05|

```

```
| 1239610373000|Wf1svKIg0qfIE5KwR...|TJHd1VBspNcua0XPn...| corriere.it|
Mozilla/5.0 (Maci...|fj2L1ILTFGMfhdr3...| 75.117.56.155|2009-04-13-08-05|
| 1239610768000|4MJR0XxiVCU1ueXKV...|10hGwmbvKf8ajoU8a...| corriere.it|
Mozilla/5.0 (Maci...|fj2L1ILTFGMfhdr3...| 75.117.56.155|2009-04-13-08-05|
| 1239610832000|gWIrpDiN57i3sHatv...|RNL4C7xPi3tdar2Uc...| corriere.it|
Mozilla/5.0 (Maci...|fj2L1ILTFGMfhdr3...| 75.117.56.155|2009-04-13-08-05|
| 1239610789000|pTne9k62kJ14QViXI...|RVxJVIQousjxUVI3r...| pixnet.net|
Mozilla/5.0 (Maci...|1bG0KiBD2xmui90kF...| 33.176.101.80|2009-04-13-08-05|
+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+
only showing top 20 rows
```

```
scala>
```

## Beispiel: Erstellen einer HCatalog-Tabelle und Ausführen eines Schreibvorgangs mit Pig

Sie können eine HCatalog-Tabelle erstellen und mit Apache Pig CatStorer über H mithilfe einer Datenquelle in Amazon S3 in sie schreiben. Für HCatalog ist erforderlich, dass Sie die Direct Write-Funktion deaktivieren. Andernfalls schlägt die Operation automatisch fehl. Legen Sie sowohl die Konfiguration `mapred.output.direct.NativeS3FileSystem` als auch `mapred.output.direct.EmrFileSystem` entweder mit der `false`-Klassifizierung oder manuell innerhalb der Grunt-Shell auf `mapred-site` fest. Das folgende Beispiel zeigt eine Tabelle, die mit der HCat CLI erstellt wurde, gefolgt von Befehlen, die in der Grunt-Shell ausgeführt wurden, um die Tabelle aus einer Beispieldatendatei in Amazon S3 zu füllen.

Um dieses Beispiel auszuführen, müssen Sie eine [Verbindung mit dem Hauptknoten über SSH herstellen](#).

Erstellen Sie eine HCatalog-Skriptdatei, `wikicount.q`, mit dem folgenden Inhalt, der eine HCatalog-Tabelle namens `wikicount` erstellt.

```
CREATE EXTERNAL TABLE IF NOT EXISTS wikicount(
  col1 string,
  col2 bigint
)
ROW FORMAT DELIMITED FIELDS TERMINATED BY '\001'
STORED AS ORC
LOCATION 's3://MyBucket/hcat/wikicount';
```

Führen Sie das Skript mit einem HCat CLI-Befehl in der Datei aus.

```
hcat -f wikicount.q
```

Im nächsten Schritt starten Sie die Grund-Shell mit der Option `-useHCatalog`, legen Konfigurationen zum Deaktivieren der Direct Write-Funktion fest, laden Daten von einem S3-Speicherort und schreiben die Ergebnisse in die Wikicount-Tabelle.

```
pig -useHCatalog
SET mapred.output.direct.NativeS3FileSystem false;
SET mapred.output.direct.EmrFileSystem false;
A = LOAD 's3://support.elasticmapreduce/training/datasets/wikistats_tiny/' USING
  PigStorage(' ') AS (Site:chararray, page:chararray, views:int, total_bytes:long);
B = GROUP A BY Site;
C = FOREACH B GENERATE group as col1, COUNT(A) as col2;
STORE C INTO 'wikicount' USING org.apache.hive.hcatalog.pig.HCatStorer();
```

## Versionsverlauf von HCatalog

In der folgenden Tabelle sind die Version von HCatalog aufgeführt, die in jeder Release-Version von Amazon EMR enthalten ist, zusammen mit den Komponenten, die mit der Anwendung installiert wurden. Informationen zu den Komponentenversionen in den einzelnen Versionen finden Sie im Abschnitt [Komponentenversion für Ihre Version in Amazon-EMR-7.x-Versionen](#), [Amazon-EMR-6.x-Versionen](#) oder [Amazon-EMR-5.x-Versionen](#).

Informationen zur HCatalog-Version

Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-5.36.2	2.3.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop-



Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
		yarn-resourcemanager, hadoop-yarn-timeline-server , hcatalog-client, hcatalog- server, hcatalog-webhcat-s erver, hive-client, mariadb-s erver
emr-7.1.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server , hcatalog-client, hcatalog- server, hcatalog-webhcat-s erver, hive-client, mariadb-s erver
emr-7.0.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server , hcatalog-client, hcatalog- server, hcatalog-webhcat-s erver, hive-client, mariadb-s erver

Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-6.15.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server , hcatalog-client, hcatalog- server, hcatalog-webhcat-s erver, hive-client, mariadb-s erver
emr-6.14.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server , hcatalog-client, hcatalog- server, hcatalog-webhcat-s erver, hive-client, mariadb-s erver

Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-6.13.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-6.12.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-6.11.1	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-6.11.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-6.10.1	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-6.10.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-6.9.1	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-6.9.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-6.8.1	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-6.8.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-6.7.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-5.36.1	2.3.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server



Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-5.36.0	2.3.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-6.6.0	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-5.35.0	2.3.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-6.5.0	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-6.4.0	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-6.3.1	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-6.3.0	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-6.2.1	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-6.2.0	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-6.1.1	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-6.1.0	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-6.0.1	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-6.0.0	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server , hcatalog-client, hcatalog- server, hcatalog-webhcat-s erver, hive-client, mariadb-s erver
emr-5.34.0	2.3.8	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server , hcatalog-client, hcatalog- server, hcatalog-webhcat-s erver, hive-client, mariadb-s erver

Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-5.33.1	2.3.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-5.33.0	2.3.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server



Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-5.32.1	2.3.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-5.32.0	2.3.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-5.31.1	2.3.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-5.31.0	2.3.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-5.30.2	2.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-5.30.1	2.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-5.30.0	2.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-5.29.0	2.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-5.28.1	2.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server , hcatalog-client, hcatalog- server, hcatalog-webhcat-s erver, hive-client, mysql-server
emr-5.28.0	2.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server , hcatalog-client, hcatalog- server, hcatalog-webhcat-s erver, hive-client, mysql-server

Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-5.27.1	2.3.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server , hcatalog-client, hcatalog- server, hcatalog-webhcat-s erver, hive-client, mysql-server
emr-5.27.0	2.3.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server , hcatalog-client, hcatalog- server, hcatalog-webhcat-s erver, hive-client, mysql-server

Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-5.26.0	2.3.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server , hcatalog-client, hcatalog- server, hcatalog-webhcat-s erver, hive-client, mysql-server
emr-5.25.0	2.3.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server , hcatalog-client, hcatalog- server, hcatalog-webhcat-s erver, hive-client, mysql-server

Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-5.24.1	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server , hcatalog-client, hcatalog- server, hcatalog-webhcat-s erver, hive-client, mysql-server
emr-5.24.0	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server , hcatalog-client, hcatalog- server, hcatalog-webhcat-s erver, hive-client, mysql-server



Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-5.23.1	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server , hcatalog-client, hcatalog- server, hcatalog-webhcat-s erver, hive-client, mysql-server
emr-5.23.0	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server , hcatalog-client, hcatalog- server, hcatalog-webhcat-s erver, hive-client, mysql-server

Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-5.22.0	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server , hcatalog-client, hcatalog- server, hcatalog-webhcat-s erver, hive-client, mysql-server
emr-5.21.2	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server , hcatalog-client, hcatalog- server, hcatalog-webhcat-s erver, hive-client, mysql-server

Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-5.21.1	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server , hcatalog-client, hcatalog- server, hcatalog-webhcat-s erver, hive-client, mysql-server
emr-5.21.0	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server , hcatalog-client, hcatalog- server, hcatalog-webhcat-s erver, hive-client, mysql-server

Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-5.20.1	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.20.0	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-5.19.1	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server , hcatalog-client, hcatalog- server, hcatalog-webhcat-s erver, hive-client, mysql-server
emr-5.19.0	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server , hcatalog-client, hcatalog- server, hcatalog-webhcat-s erver, hive-client, mysql-server

Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-5.18.1	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server , hcatalog-client, hcatalog- server, hcatalog-webhcat-s erver, hive-client, mysql-server
emr-5.18.0	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server , hcatalog-client, hcatalog- server, hcatalog-webhcat-s erver, hive-client, mysql-server

Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-5.17.2	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.17.1	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-5.17.0	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.16.1	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server



Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-5.16.0	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.15.1	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-5.15.0	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server , hcatalog-client, hcatalog- server, hcatalog-webhcat-s erver, hive-client, mysql-server
emr-5.14.2	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server , hcatalog-client, hcatalog- server, hcatalog-webhcat-s erver, hive-client, mysql-server

Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-5.14.1	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server , hcatalog-client, hcatalog- server, hcatalog-webhcat-s erver, hive-client, mysql-server
emr-5.14.0	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server , hcatalog-client, hcatalog- server, hcatalog-webhcat-s erver, hive-client, mysql-server

Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-5.13.1	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server , hcatalog-client, hcatalog- server, hcatalog-webhcat-s erver, hive-client, mysql-server
emr-5.13.0	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server , hcatalog-client, hcatalog- server, hcatalog-webhcat-s erver, hive-client, mysql-server

Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-5.12.3	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server , hcatalog-client, hcatalog- server, hcatalog-webhcat-s erver, hive-client, mysql-server
emr-5.12.2	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server , hcatalog-client, hcatalog- server, hcatalog-webhcat-s erver, hive-client, mysql-server

Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-5.12.1	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.12.0	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-5.11.4	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server , hcatalog-client, hcatalog- server, hcatalog-webhcat-s erver, hive-client, mysql-server
emr-5.11.3	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server , hcatalog-client, hcatalog- server, hcatalog-webhcat-s erver, hive-client, mysql-server

Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-5.11.2	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server , hcatalog-client, hcatalog- server, hcatalog-webhcat-s erver, hive-client, mysql-server
emr-5.11.1	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server , hcatalog-client, hcatalog- server, hcatalog-webhcat-s erver, hive-client, mysql-server



Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-5.11.0	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server , hcatalog-client, hcatalog- server, hcatalog-webhcat-s erver, hive-client, mysql-server
emr-5.10.1	2.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server , hcatalog-client, hcatalog- server, hcatalog-webhcat-s erver, hive-client, mysql-server

Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-5.10.0	2.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.9.1	2.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-5.9.0	2.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.8.3	2.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-5.8.2	2.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.8.1	2.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-5.8.0	2.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server , hcatalog-client, hcatalog- server, hcatalog-webhcat-s erver, hive-client, mysql-server
emr-5.7.1	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server , hcatalog-client, hcatalog- server, hcatalog-webhcat-s erver, hive-client, mysql-server

Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-5.7.0	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server , hcatalog-client, hcatalog- server, hcatalog-webhcat-s erver, hive-client, mysql-server
emr-5.6.1	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server , hcatalog-client, hcatalog- server, hcatalog-webhcat-s erver, hive-client, mysql-server

Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-5.6.0	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server , hcatalog-client, hcatalog- server, hcatalog-webhcat-s erver, hive-client, mysql-server
emr-5.5.4	2.1.1	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-mapred, hadoop- hdfs-datanode, hadoop- hdfs-library, hadoop-hdfs- namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hcatalog-client, hcatalog-server, hcatalog- webhcat-server, hive-client, mysql-server

Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-5.5.3	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.5.2	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server



Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-5.5.1	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.5.0	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-5.4.1	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.4.0	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-5.3.2	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.3.1	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-5.3.0	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.2.3	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-5.2.2	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.2.1	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-5.2.0	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.1.1	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-5.1.0	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.0.3	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-5.0.2	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.0.1	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server



Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-5.0.0	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-4.9.6	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-4.9.5	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-4.9.4	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-4.9.3	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-4.9.2	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-4.9.1	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-4.8.5	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-4.8.4	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-4.8.3	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-4.8.2	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-4.8.1	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-4.8.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-4.7.4	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-4.7.3	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-4.7.2	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server



Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-4.7.1	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-4.7.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-4.6.1	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hd fs-datanode, hadoop-hdfs- library, hadoop-hdfs-nameno de, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hcatalog-client, hcatalog-server, hcatalog- webhcat-server, hive-client, hive-metastore-server, mysql- server
emr-4.6.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hd fs-datanode, hadoop-hdfs- library, hadoop-hdfs-nameno de, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hcatalog-client, hcatalog-server, hcatalog- webhcat-server, hive-client, hive-metastore-server, mysql- server

Amazon-EMR-Version sbezeichnung	HCatalog-Version	Mit HCatalog installierte Komponenten
emr-4.5.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hd fs-datanode, hadoop-hdfs- library, hadoop-hdfs-nameno de, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hcatalog-client, hcatalog-server, hcatalog- webhcat-server, hive-client, hive-metastore-server, mysql- server
emr-4.4.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hd fs-datanode, hadoop-hdfs- library, hadoop-hdfs-nameno de, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hcatalog-client, hcatalog-server, hcatalog- webhcat-server, hive-client, hive-metastore-server, mysql- server

# Apache Hive

Bei Hive handelt es sich um ein Open-Source-Data-Warehouse und Analyse-Paket von Datensätzen, das auf einem Hadoop-Cluster ausgeführt wird. Für Hive-Skripte wird die SQL-ähnliche Sprache Hive QL (Query Language) verwendet, die Programmiermodelle abstrahiert und typische Data Warehouse-Interaktionen unterstützt. Mit Hive können Sie die Komplexität vermeiden, die mit dem Schreiben von Tez-Jobs verbunden ist, die auf gerichteten azyklischen Graphen (DAGs) oder MapReduce Programmen in einer Computersprache niedrigerer Stufe wie Java basieren.

Hive erweitert das SQL-Paradigma durch Einschließen von Serialisierungsformaten. Sie können die Abfrageverarbeitung auch anpassen, indem Sie ein Tabellenschema erstellen, das mit Ihren Daten übereinstimmt, ohne die eigentlichen Daten zu bearbeiten. Während SQL nur primitive Wertetypen (z. B. Datumsangaben, Zahlen und Zeichenfolgen) unterstützt, handelt es sich bei den Werten in Hive-Tabellen um strukturierte Elemente wie beispielsweise JSON-Objekte, benutzerdefinierte Datentypen oder Java-Funktionen.

Weitere Informationen zu Hive finden Sie unter <http://hive.apache.org/>.

Die folgende Tabelle listet die Version von Hive auf, die in der neuesten Version der Amazon-EMR-7.x-Serie enthalten ist, zusammen mit den Komponenten, die Amazon EMR mit Hive installiert.

[Informationen zur Version der Komponenten, die in dieser Version mit Hive installiert wurden, finden Sie unter Komponentenversionen von Version 7.1.0.](#)

Hive-Versionsinformationen für emr-7.1.0

Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-7.1.0	Hive 3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-ya

Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
		rn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, tez-on-worker, zookeeper-client, zookeeper-server

Die folgende Tabelle listet die Version von Hive auf, die in der neuesten Version der Amazon-EMR-6.x-Serie enthalten ist, zusammen mit den Komponenten, die Amazon EMR mit Hive installiert.

Die Version der Komponenten, die mit Hive in dieser Version installiert wurden, finden Sie unter [Komponentenversionen der Version 6.15.0](#).

Hive-Versionsinformationen für emr-6.15.0

Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-6.15.0	Hive 3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, tez-on-worker, zookeeper-client, zookeeper-server

Die folgende Tabelle listet die Version von Hive auf, die in der neuesten Version der Amazon-EMR-5.x-Serie enthalten ist, zusammen mit den Komponenten, die Amazon EMR mit Hive installiert.

[Informationen zur Version der Komponenten, die in dieser Version mit Hive installiert wurden, finden Sie unter Komponentenversionen von Version 5.36.2.](#)

Hive-Versionsinformationen für emr-5.36.2

Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-5.36.2	Hive 2.3.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn

Ab Amazon EMR 5.18.0 können Sie das Amazon-EMR-Artefakt-Repository verwenden, um Ihren Auftragscode anhand der genauen Versionen von Bibliotheken und Abhängigkeiten zu erstellen, die mit bestimmten Amazon-EMR-Versionen verfügbar sind. Weitere Informationen finden Sie unter [Überprüfen von Abhängigkeiten mithilfe des Amazon-EMR-Artefakt-Repositorys](#).

Themen

- [Unterschiede und Überlegungen für Hive auf Amazon EMR](#)
- [Konfigurieren eines externen Metastores für Hive](#)
- [Den Hive-JDBC-Treiber verwenden](#)
- [Verbessern der Hive-Leistung](#)
- [Verwendung von Hive Live Long and Process \(LLAP\)](#)

- [Verschlüsselung im Hive](#)
- [Versionsverlauf von Hive](#)

## Unterschiede und Überlegungen für Hive auf Amazon EMR

### Unterschiede zwischen Apache Hive in Amazon EMR und Apache Hive

In diesem Abschnitt werden die Unterschiede zwischen Hive in Amazon EMR und den unter <http://svn.apache.org/viewvc/hive/branches/> verfügbaren Hive-Standardversionen beschrieben.

#### Hive-Autorisierung

Amazon EMR unterstützt [Hive-Autorisierung](#) für HDFS, aber nicht für EMRFS und Amazon S3. Amazon-EMR-Cluster werden standardmäßig mit deaktivierter Autorisierung ausgeführt.

#### Dateizusammenführung mit Hive in Amazon S3

Apache Hive führt kleine Dateien am Ende eines Map-Auftrags zusammen, wenn `hive.merge.mapfiles` auf "true" gesetzt ist und die Zusammenführung nur dann ausgelöst wird, wenn die durchschnittliche Ausgabegröße des Auftrags kleiner als die Einstellung `hive.merge.smallfiles.avgsize` ist. Amazon EMR; Hive verhält sich genau gleich, wenn sich der Endausgabepfad in HDFS befindet. Wenn sich der Ausgabepfad in Amazon S3 befindet, wird der Parameter `hive.merge.smallfiles.avgsize` jedoch ignoriert. In diesem Fall wird die Zusammenführungsaufgabe immer dann ausgelöst, wenn `hive.merge.mapfiles` auf true gesetzt ist.

#### ACID-Transaktionen und Amazon S3

Amazon EMR 6.1.0 und höher unterstützt Hive-ACID-Transaktionen (Atomicity, Consistency, Isolation, Durability), sodass sie den ACID-Eigenschaften einer Datenbank entsprechen. Mit diesem Feature können Sie INSERT-, UPDATE-, DELETE- und MERGE-Operationen in Hive-Tabellen mit Daten in Amazon Simple Storage Service (Amazon S3) ausführen.

#### Hive Live Long and Process (LLAP)

Die [LLAP-Funktionalität](#), die in Version 2.0 von Standard-Apache Hive hinzugefügt wurde, wird in Hive 2.1.0 für Amazon-EMR-Version 5.0 nicht unterstützt.

Amazon-EMR-Version 6.0.0 und höher unterstützt die Live Long and Process (LLAP)-Funktionalität für Hive. Weitere Informationen finden Sie unter [Verwenden von Hive LLAP](#).

## Hive-Unterschiede zwischen Amazon-EMR-Version 4.x und 5.x

In diesem Abschnitt werden die Unterschiede behandelt, die Sie berücksichtigen sollten, bevor Sie eine Hive-Implementierung von Hive Version 1.0.0 in Amazon-EMR-Version 4.x zu Hive 2.x in Amazon-EMR-Version 5.x migrieren.

### Betriebliche Unterschiede und Überlegungen

- Die Unterstützung für [ACID-Transaktionen \(Atomizität, Konsistenz, Isolation und Dauerhaftigkeit\)](#) wurde ergänzt: Dieser Unterschied zwischen Hive 1.0.0 für Amazon EMR 4.x und Standard-Apache-Hive wurde entfernt.
- Direkte Schreibvorgänge in Amazon S3 beseitigt: Dieser Unterschied zwischen Hive 1.0.0 auf Amazon EMR und dem Standard-Apache-Hive wurde beseitigt. Hive 2.1.0 in Amazon-EMR-Version 5.x erstellt, liest und schreibt jetzt aus und zu temporären Dateien, die in Amazon S3 gespeichert werden. Daher müssen Sie bei Lese- und Schreibvorgängen für dieselbe Tabelle keine temporäre Tabelle mehr im lokalen HDFS-Dateisystem des Clusters erstellen, um das Problem zu umgehen. Wenn Sie versionierte Buckets verwenden, stellen Sie sicher, dass diese temporären Dateien wie unten beschrieben verwaltet werden.
- Verwalten Sie temporäre Dateien, wenn Sie versionierte Amazon-S3-Buckets verwenden: Wenn Sie Hive-Abfragen ausführen, bei denen das Ziel der generierten Daten Amazon S3 ist, werden viele temporäre Dateien und Verzeichnisse erstellt. Hierbei handelt es sich, wie bereits beschrieben, um ein neues Verhalten. Wenn Sie versionierte S3-Buckets verwenden, wird Amazon S3 von diesen temporären Dateien überfrachtet und es fallen Kosten an, wenn sie nicht gelöscht werden. Passen Sie Ihre Lebenszyklusregeln an, damit Daten mit einem `/_tmp`-Präfix nach einem kurzen Zeitraum, wie z. B. nach fünf Tagen, gelöscht werden. Weitere Informationen finden Sie unter [Angabe einer Lebenszykluskonfiguration](#).
- Aktualisierung von Log4j auf log4j 2: Wenn Sie log4j verwenden, müssen Sie aufgrund dieser Aktualisierung möglicherweise Ihre Protokollierungskonfiguration ändern. Weitere Informationen finden Sie unter [Apache log4j 2](#).

### Unterschiede und Überlegungen im Bezug auf Leistung

- Leistungsunterschiede zu Tez: Mit Amazon EMR Version 5.x ist Tez die Standard-Ausführungse-engine für Hive anstelle von. MapReduce Tez bietet eine verbesserte Leistung für die meisten Workflows.



- Tabellen mit zahlreichen Partitionen: Abfragen, die eine große Anzahl dynamischer Partitionen generieren, schlagen möglicherweise fehl. Außerdem kann die Ausführung von Abfragen, die aus Tabellen mit vielen Partitionen auswählen, länger dauern als erwartet. Beispielsweise kann eine Auswahl aus 100.000 Partitionen 10 Minuten oder länger dauern.

## Zusätzliche Features von Hive in Amazon EMR

Amazon EMR erweitert Hive um neue Funktionen, die die Hive-Integration mit anderen AWS Diensten unterstützen, z. B. die Fähigkeit, von Amazon Simple Storage Service (Amazon S3) und DynamoDB zu lesen und in diese zu schreiben.

### Variablen in Hive

Sie können Variablen in Ihre Skripts aufnehmen, indem Sie das Dollarzeichen und geschweifte Klammern verwenden.

```
add jar ${LIB}/jsonserde.jar
```

Sie übergeben die Werte dieser Variablen an Hive in der Befehlszeile mit dem Parameter `-d`, wie im folgenden Beispiel dargestellt:

```
-d LIB=s3://elasticmapreduce/samples/hive-ads/lib
```

Sie können die Werte auch an Schritte übergeben, die Hive-Skripts ausführen.

So übergeben Sie Variablenwerte mithilfe der Konsole an Hive-Schritte

1. Öffnen Sie die Amazon EMR-Konsole unter <https://console.aws.amazon.com/emr>.
2. Wählen Sie Cluster erstellen.
3. Wählen Sie im Abschnitt Steps (Schritte) in Add Step (Schritt hinzufügen) die Option Hive Program (Hive-Programm) in der Liste und Configure and add (Konfigurieren und hinzufügen) aus.
4. Geben Sie im Dialogfeld Add Step (Schritt hinzufügen) die Parameter mithilfe der folgenden Tabelle als Leitfaden an. Wählen Sie anschließend Add (Hinzufügen) aus.


Feld	Aktion
Skript-S3-Speicherort*	Geben Sie den URI an, an dem das Skript in Amazon S3 gespeichert ist. Der Wert muss das Format <i>BucketName /path/ScriptName</i> haben. Zum Beispiel: <code>s3://elasticmapreduce/samples/hive-ads/libs/response-time-stats.q</code> .
S3-Eingabespeicherort	Geben Sie wahlweise den URI an, an dem sich Ihre Eingabedateien in Amazon S3 befinden. Der Wert muss das Format <i>BucketName /Pfad</i> /haben. Wenn der Wert angegeben ist, wird er an das Hive-Skript als Parameter mit dem Namen INPUT übergeben. Zum Beispiel: <code>s3://elasticmapreduce/samples/hive-ads/tables/</code> .
S3-Ausgabespeicherort	Geben Sie wahlweise den URI an, an dem die Ausgabe in Amazon S3 gespeichert werden soll. Der Wert muss das Format <i>BucketName /Pfad</i> haben. Wenn der Wert angegeben ist, wird er an das Hive-Skript als Parameter mit dem Namen OUTPUT übergeben. Zum Beispiel: <code>s3://mybucket/hive-ads/output/</code> .
Argumente	<p>Wahlweise können Sie eine Liste von Argumenten (durch Leerzeichen getrennte Zeichenfolgen) zur Übergabe an Hive eingeben. Wenn Sie eine Pfadvariable in Ihrem Hive-Skript mit dem Namen <code>_\${SAMPLE}</code> definiert haben, zum Beispiel:</p> <pre data-bbox="451 1241 1507 1476">CREATE EXTERNAL TABLE logs (requestBeginTime STRING, requestEndTime STRING, hostname STRING) PARTITIONED BY (dt STRING) \ ROW FORMAT serde 'com.amazon.elasticmapreduce.JsonSerde' WITH SERDEPROPERTIES ( 'paths'='requestBeginTime, requestEndTime, hostname' ) LOCATION '\$_{SAMPLE}/tables/impressions';</pre> <p>Um einen Wert für die Variable zu übergeben, geben Sie im Fenster Arguments (Argumente) Folgendes ein:</p> <pre data-bbox="451 1640 1507 1682">-d SAMPLE=s3://elasticmapreduce/samples/hive-ads/ .</pre>

Feld	Aktion
Action on Failure	<p>Hiermit wird bestimmt, welche Schritte der Cluster als Reaktion auf Fehler ausführt. Die möglichen Werte für diese Einstellung sind folgende:</p> <ul style="list-style-type: none"> <li>• <b>Terminate cluster (Cluster beenden):</b> Wenn der Schritt fehlschlägt, wird der Cluster beendet. Wenn für den Cluster der Beendigungsschutz aktiviert UND Keepalive aktiviert ist, wird er nicht beendet.</li> <li>• <b>Cancel and wait (Abbrechen und warten):</b> Wenn der Schritt fehlschlägt, werden die verbleibenden Schritte abgebrochen. Wenn für den Cluster Keepalive aktiviert ist, wird der Cluster nicht beendet.</li> <li>• <b>Continue (Fortfahren):</b> Wenn der Schritt fehlschlägt, wird mit dem nächsten Schritt fortgefahren.</li> </ul>

5. Wählen Sie die Werte wie notwendig und anschließend Create cluster (Cluster erstellen) aus.

Um Variablenwerte an Hive zu übergeben, verwenden Sie die AWS CLI

Um Variablenwerte mithilfe von an Hive-Schritte zu übergeben AWS CLI, verwenden Sie den `--steps` Parameter und fügen Sie eine Argumentliste hinzu.

 Note

Linux-Zeilenfortsetzungszeichen (`\`) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (`^`).

```
aws emr create-cluster --name "Test cluster" --release-label emr-7.1.0 \
--applications Name=Hive Name=Pig --use-default-roles --ec2-attributes
  KeyName=myKey --instance-type m5.xlarge --instance-count 3 \
--steps Type=Hive,Name="Hive Program",ActionOnFailure=CONTINUE,Args=[-f,s3://
elasticmapreduce/samples/hive-ads/libs/response-time-stats.q,-d,INPUT=s3://
elasticmapreduce/samples/hive-ads/tables,-d,OUTPUT=s3://mybucket/hive-ads/output/,
-d,SAMPLE=s3://elasticmapreduce/samples/hive-ads/]
```

Weitere Informationen zur Verwendung von Amazon EMR-Befehlen finden Sie AWS CLI unter <https://docs.aws.amazon.com/cli/latest/reference/emr>.

## So übergeben Sie Variablenwerte mithilfe des Java SDK an Hive-Schritte

- Im folgenden Beispiel wird gezeigt, wie Variablen mithilfe des SDK an Schritte übergeben werden. Weitere Informationen finden Sie unter [Class StepFactory](#) in der AWS SDK for Java API-Referenz.

```
StepFactory stepFactory = new StepFactory();

StepConfig runHive = new StepConfig()
    .withName("Run Hive Script")
    .withActionOnFailure("TERMINATE_JOB_FLOW")
    .withHadoopJarStep(stepFactory.newRunHiveScriptStep("s3://mybucket/script.q",
        Lists.newArrayList("-d", "LIB= s3://elasticmapreduce/samples/hive-ads/lib")));
```

## Amazon-EMR-Hive-Abfragen zur Unterstützung partieller DynamoDB-Schemas

Amazon EMR Hive bietet maximale Flexibilität beim Abfragen von DynamoDB-Tabellen, da Sie eine Teilmenge von Spalten angeben können, nach denen Sie Daten filtern können, und nicht alle Spalten in die Abfrage einschließen müssen. Diese Technik der Teilschemaabfrage ist effektiv, wenn Sie ein Sparse-Datenbankschema vorliegen haben und Datensätze nach wenigen Spalten filtern möchten, z. B. nach Zeitstempel.

Im folgenden Beispiel wird gezeigt, wie Sie eine Hive-Abfrage für folgende Aufgaben verwenden:

- Erstellen Sie eine DynamoDB-Tabelle.
- Wählen Sie eine Teilmenge von Elementen (Zeilen) in DynamoDB aus und grenzen Sie die Daten basierend auf bestimmten Spalten weiter ein.
- Kopieren Sie die Ergebnisdaten in Amazon S3.

```
DROP TABLE dynamodb;
DROP TABLE s3;

CREATE EXTERNAL TABLE dynamodb(hashKey STRING, recordTimeStamp BIGINT, fullColumn
map<String, String>)
    STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
    TBLPROPERTIES (
        "dynamodb.table.name" = "myTable",
        "dynamodb.throughput.read.percent" = ".1000",
        "dynamodb.column.mapping" = "hashKey:HashKey,recordTimeStamp:RangeKey");
```

```
CREATE EXTERNAL TABLE s3(map<String, String>)
  ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
  LOCATION 's3://bucketname/path/subpath/';

INSERT OVERWRITE TABLE s3 SELECT item fullColumn FROM dynamodb WHERE recordTimeStamp <
"2012-01-01";
```

Die folgende Tabelle zeigt die Abfragesyntax für die Auswahl einer beliebigen Kombination von Elementen aus DynamoDB.

Abfragebeispiel:	Ergebnisbeschreibung
<code>SELECT * FROM <i>table_name</i> ;</code>	Wählt alle Elemente (Zeilen) in einer bestimmten Tabelle aus und bezieht Daten aus allen für diese Elemente verfügbaren Spalten ein.
<code>SELECT * FROM <i>table_name</i> WHERE <i>field_name</i> =<i>value</i>;</code>	Wählt einige Elemente (Zeilen) in einer bestimmten Tabelle aus und bezieht Daten aus allen für diese Elemente verfügbaren Spalten ein.
<code>SELECT <i>column1_name</i> , <i>column2_name</i> , <i>column3_name</i> FROM <i>table_name</i> ;</code>	Wählt alle Elemente (Zeilen) in einer bestimmten Tabelle aus und bezieht Daten aus einigen für diese Elemente verfügbaren Spalten ein.
<code>SELECT <i>column1_name</i> , <i>column2_name</i> , <i>column3_name</i> FROM <i>table_name</i> WHERE <i>field_name</i> =<i>value</i>;</code>	Wählt einige Elemente (Zeilen) in einer bestimmten Tabelle aus und bezieht Daten aus einigen für diese Elemente verfügbaren Spalten ein.

## Kopieren von Daten zwischen DynamoDB-Tabellen in verschiedenen AWS -Regionen

Amazon EMR Hive stellt die `dynamodb.region`-Eigenschaft bereit, die Sie pro DynamoDB-Tabelle festlegen können. Wenn `dynamodb.region` in zwei Tabellen unterschiedlich festgelegt ist, erscheinen alle Daten, die Sie zwischen den Tabellen kopieren, automatisch in den angegebenen Regionen.

Das folgende Beispiel zeigt das Erstellen einer DynamoDB-Tabelle mit einem Hive-Skript, das die `dynamodb.region`-Eigenschaft festlegt:

**Note**

Die tabellenspezifischen Regionseigenschaften überschreiben die globalen Hive-Eigenschaften.

```
CREATE EXTERNAL TABLE dynamodb(hashKey STRING, recordTimeStamp BIGINT, map<String, String> fullColumn)
  STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
  TBLPROPERTIES (
    "dynamodb.table.name" = "myTable",
    "dynamodb.region" = "eu-west-1",
    "dynamodb.throughput.read.percent" = ".1000",
    "dynamodb.column.mapping" = "hashKey:HashKey,recordTimeStamp:RangeKey");
```

## Festlegen von DynamoDB-Durchsatzwerten pro Tabelle

Amazon EMR Hive ermöglicht es Ihnen, DynamoDB `readThroughputPercent` und `writeThroughputPercent` Einstellungen pro Tabelle in der Tabellendefinition festzulegen. Das folgende Amazon-EMR-Hive-Skript zeigt, wie die Durchsatzwerte festgelegt werden. Weitere Informationen zu DynamoDB-Durchsatzwerten finden Sie unter [Angeben von Lese- und Schreibanforderungen für Tabellen](#).

```
CREATE EXTERNAL TABLE dynamodb(hashKey STRING, recordTimeStamp BIGINT, map<String, String> fullColumn)
  STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
  TBLPROPERTIES (
    "dynamodb.table.name" = "myTable",
    "dynamodb.throughput.read.percent" = ".4",
    "dynamodb.throughput.write.percent" = "1.0",
    "dynamodb.column.mapping" = "hashKey:HashKey,recordTimeStamp:RangeKey");
```

## Konfigurieren eines externen Metastores für Hive

Hive zeichnet Metastore-Daten standardmäßig in einer MySQL-Datenbank im Dateisystem des Primärknotens auf. Die Metastore enthält eine Beschreibung der Tabelle und der zugrunde liegenden Daten, auf denen sie beruht. Das schließt auch die Partitionsnamen, Datentypen usw. ein. Wenn ein Cluster beendet wird, werden alle Cluster-Knoten heruntergefahren – einschließlich Primärknoten.

Wenn dies geschieht, gehen lokale Daten verloren, da die Dateisysteme der Knoten flüchtigen Speicher verwenden. Wenn Sie der Metastore dauerhaft sein soll, müssen Sie einen externen Metastore außerhalb des Clusters erstellen.

Ihnen stehen zwei Optionen für einen externen Metastore zur Verfügung:

- AWS Glue-Datenkatalog (nur Amazon EMR Version 5.8.0 oder höher).

Weitere Informationen finden Sie unter [Verwenden des AWS Glue-Datenkatalogs als Metastore für Hive](#).

- Amazon RDS oder Amazon Aurora.

Weitere Informationen finden Sie unter [Verwenden einer externen MySQL-Datenbank oder von Amazon Aurora](#).

#### Note

Wenn Sie Hive 3 verwenden und zu viele Verbindungen zu Hive Metastore haben, konfigurieren Sie den Parameter `datanucleus.connectionPool.maxPoolSize` so, dass er einen kleineren Wert hat, oder erhöhen Sie die Anzahl der Verbindungen, die der Datenbankserver verarbeiten kann. Die erhöhte Anzahl von Verbindungen ist darauf zurückzuführen, wie Hive die maximale Anzahl von JDBC-Verbindungen berechnet. [Informationen zur Berechnung des optimalen Leistungswerts finden Sie unter Hive-Konfigurationseigenschaften](#).

## Verwenden des AWS Glue-Datenkatalogs als Metastore für Hive

Mit Amazon EMR Version 5.8.0 oder höher können Sie Hive so konfigurieren, dass der AWS Glue-Datenkatalog als Metastore verwendet wird. Wir empfehlen diese Konfiguration, wenn Sie einen persistenten Metastore oder einen gemeinsamen Metastore für verschiedene Cluster, Services, Anwendungen oder AWS -Konten benötigen.

AWS Glue ist ein vollständig verwalteter ETL-Service (Extrahieren, Transformieren und Laden), mit dem Sie Ihre Daten einfach und kostengünstig kategorisieren, bereinigen, anreichern und zuverlässig zwischen verschiedenen Datenspeichern verschieben können. Der AWS Glue Data Catalog bietet ein einheitliches Metadaten-Repository für eine Vielzahl von Datenquellen und Datenformaten, das sich in Amazon EMR sowie Amazon RDS, Amazon Redshift, Redshift Spectrum, Athena und jede

mit dem Apache Hive Metastore kompatible Anwendung integrieren lässt. AWS Glue-Crawler können automatisch Schemas aus Quelldaten in Amazon S3 ableiten und die zugehörigen Metadaten im Datenkatalog speichern. Weitere Informationen zum Datenkatalog finden Sie unter [Füllen des AWS Glue-Datenkatalogs im AWS Glue-Entwicklerhandbuch](#).

Für AWS Glue fallen separate Gebühren an. Es gibt eine monatliche Gebühr für das Speichern und Zugreifen auf die Metadaten im Datenkatalog, einen Stundensatz, der pro Minute für AWS Glue ETL-Jobs und Crawler-Laufzeit abgerechnet wird, und einen Stundensatz, der pro Minute für jeden bereitgestellten Entwicklungsendpunkt abgerechnet wird. Der Datenkatalog ermöglicht Ihnen die gebührenfreie Speicherung von bis zu einer Million Objekten. Wenn Sie mehr als eine Million Objekte speichern, fällt eine Gebühr von 1 USD für jeweils 100.000 zusätzliche Objekte an. Im Data Catalog ist ein Objekt eine Tabelle, eine Partition oder eine Datenbank. Weitere Informationen finden Sie unter [Glue-Preise](#).

#### Important

Wenn Sie vor dem 14. August 2017 Tabellen mit Amazon Athena oder Amazon Redshift Spectrum erstellt haben, werden Datenbanken und Tabellen in einem von Athena verwalteten Katalog gespeichert, der vom Glue-Datenkatalog getrennt ist. AWS Um Amazon EMR mit diesen Tabellen zu integrieren, müssen Sie ein Upgrade auf den AWS Glue Data Catalog durchführen. Weitere Informationen finden Sie unter [Upgrade auf den AWS Glue-Datenkatalog](#) im Amazon Athena Athena-Benutzerhandbuch.

## AWS Glue Data Catalog als Metastore angeben

Sie können den AWS Glue-Datenkatalog mithilfe der Amazon EMR-API, der AWS Management Console, der AWS CLI, oder als Metastore angeben. Wenn Sie die CLI oder API verwenden, verwenden Sie die Konfigurationsklassifizierung für Hive, um den Data Catalog anzugeben. Darüber hinaus können Sie mit Amazon EMR 5.16.0 und höher die Konfigurationsklassifizierung verwenden, um einen Datenkatalog in einem anderen Format anzugeben. **AWS-Konto** Wenn Sie die Konsole verwenden, können Sie den Data Catalog mit den erweiterten Optionen oder den Schnelloptionen angeben.

### New console

Um AWS Glue Data Catalog mit der neuen Konsole als Hive-Metastore anzugeben

1. Melden Sie sich bei der AWS Management Console an und öffnen Sie die Amazon EMR-Konsole unter <https://console.aws.amazon.com/emr>.



2. Wählen Sie im linken Navigationsbereich unter EMR in EC2 die Option Cluster und dann Cluster erstellen aus.
3. Wählen Sie unter Anwendungspaket die Option Core Hadoop, HBase oder Benutzerdefiniert aus. Wenn Sie Ihren Cluster anpassen, stellen Sie sicher, dass Sie Hive oder HCatalog als eine Ihrer Anwendungen auswählen.
4. Aktivieren Sie unter Einstellungen für den AWS Glue Data Catalog das Kontrollkästchen Für Hive-Tabellenmetadaten verwenden.
5. Wählen Sie alle anderen Optionen aus, die für Ihren Cluster gelten.
6. Um Ihren Cluster jetzt zu starten, wählen Sie Cluster erstellen aus.

## Old console

Um AWS Glue Data Catalog mit der alten Konsole als Hive-Metastore zu spezifizieren

1. Navigieren Sie zur neuen Amazon-EMR-Konsole und wählen Sie in der Seitennavigation die Option Zur alten Konsole wechseln aus. Weitere Informationen darüber, was Sie erwartet, wenn Sie zur alten Konsole wechseln, finden Sie unter [Verwenden der alten Konsole](#).
2. Wählen Sie Create Cluster (Cluster erstellen) und Go to advanced options (Zu erweiterten Optionen) aus.
3. Wählen Sie für Release die Option emr-5.8.0 oder höher aus.
4. Wählen Sie unter Release (Version) entweder Hive (Hive) oder HCatalog (HCatalog) aus.
5. Aktivieren Sie unter Einstellungen für den AWS Glue Data Catalog das Kontrollkästchen Für Hive-Tabellenmetadaten verwenden.
6. Wählen Sie weitere Optionen für Ihren Cluster nach Bedarf aus, wählen Sie Next (Weiter) aus und konfigurieren Sie dann weitere Cluster-Optionen gemäß Ihrer Anwendung.

## CLI

Um den AWS Glue-Datenkatalog als Hive-Metastore zu spezifizieren mit AWS CLI

Weitere Informationen zum Angeben einer Konfigurationsklassifizierung mithilfe der AWS CLI und EMR-API finden Sie unter [Anwendungen konfigurieren](#).

- Geben Sie den Wert für `hive.metastore.client.factory.class` mittels der Konfigurationsklassifizierung `hive-site` wie im folgenden Beispiel gezeigt an:

```
[
  {
    "Classification": "hive-site",
    "Properties": {
      "hive.metastore.client.factory.class":
"com.amazonaws.glue.catalog.metastore.AWSGlueDataCatalogHiveClientFactory"
    }
  }
]
```

Wenn Sie in den EMR-Release-Versionen 5.28.0, 5.28.1, 5.29.0 oder 6.x einen Cluster mit dem AWS Glue-Datenkatalog als Metastore erstellen, setzen Sie den Wert auf `hive.metastore.schema.validation false`. Dadurch wird verhindert, dass Hive und HCatalog das Metastore-Schema gegen MySQL validieren. Ohne diese Konfiguration wird die Primär-Instance-Gruppe nach der Neukonfiguration in Hive oder HCatalog gesperrt.

```
[
  {
    "Classification": "hive-site",
    "Properties": {
      "hive.metastore.client.factory.class":
"com.amazonaws.glue.catalog.metastore.AWSGlueDataCatalogHiveClientFactory",
      "hive.metastore.schema.validation": "false"
    }
  }
]
```

Wenn Sie bereits über einen Cluster in EMR-Version 5.28.0, 5.28.1 oder 5.29.0 verfügen, können Sie die Primär-Instance-Gruppe `hive.metastore.schema.validation` mit folgenden Informationen auf `false` festlegen:

```
Classification = hive-site
Property       = hive.metastore.schema.validation
Value         = false
```

Um einen Datenkatalog in einem anderen AWS Konto anzugeben, fügen Sie die Eigenschaft hinzu, wie im folgenden Beispiel gezeigt. `hive.metastore.glue.catalogid` Ersetzen Sie *acct-id* durch das AWS -Konto des Data Catalog.

```
[
  {
    "Classification": "hive-site",
    "Properties": {
      "hive.metastore.client.factory.class":
"com.amazonaws.glue.catalog.metastore.AWSGlueDataCatalogHiveClientFactory",
      "hive.metastore.schema.validation": "false",
      "hive.metastore.glue.catalogid": "acct-id"
    }
  }
]
```

## IAM-Berechtigungen

Das EC2-Instanzprofil für einen Cluster muss über IAM-Berechtigungen für AWS Glue-Aktionen verfügen. Wenn Sie die Verschlüsselung für AWS Glue Data Catalog-Objekte aktivieren, muss die Rolle außerdem berechtigt sein, die für die Verschlüsselung AWS KMS key verwendeten Daten zu verschlüsseln, zu entschlüsseln und zu generieren.

### Berechtigungen für AWS Glue-Aktionen

Wenn Sie das standardmäßige EC2-Instance-Profil für Amazon EMR verwenden, ist keine Aktion erforderlich. Die `AmazonElasticMapReduceforEC2Role` verwaltete Richtlinie, die an die angehängt ist, `EMR_EC2_DefaultRole` ermöglicht alle erforderlichen AWS Glue-Aktionen. Wenn Sie jedoch ein benutzerdefiniertes EC2-Instanzprofil und Berechtigungen angeben, müssen Sie die entsprechenden AWS Glue-Aktionen konfigurieren. Verwenden Sie hierzu die verwaltete Richtlinie `AmazonElasticMapReduceforEC2Role` als Ausgangspunkt. Weitere Informationen finden Sie unter [Servicerolle für Cluster-EC2-Instances \(EC2-Instance-Profil\)](#) im Verwaltungshandbuch für Amazon EMR.

## Berechtigungen zum Verschlüsseln und Entschlüsseln von AWS Glue Data Catalog

Ihr Instance-Profil benötigt die Erlaubnis, Daten mithilfe Ihres Schlüssels zu verschlüsseln und zu entschlüsseln. Sie müssen diese Berechtigungen nicht konfigurieren, wenn die beiden folgenden Aussagen zutreffen:

- Sie aktivieren die Verschlüsselung für AWS Glue Data Catalog-Objekte mithilfe von verwalteten Schlüsseln für AWS Glue.
- Sie verwenden einen Cluster, der sich im selben Verzeichnis befindet AWS-Konto wie der AWS Glue-Datenkatalog.

Andernfalls müssen Sie die folgende Anweisung der Berechtigungsrichtlinie hinzufügen, die mit Ihrem EC2-Instance-Profil verknüpft ist.

```
[
  {
    "Version": "2012-10-17",
    "Statement": [
      {
        "Effect": "Allow",
        "Action": [
          "kms:Decrypt",
          "kms:Encrypt",
          "kms:GenerateDataKey"
        ],
        "Resource": "arn:aws:kms:region:acct-
id:key/12345678-1234-1234-1234-123456789012"
      }
    ]
  }
]
```

Weitere Informationen zur Verschlüsselung von AWS Glue Data Catalog finden Sie unter [Verschlüsselung Ihres Datenkatalogs](#) im AWS Glue Developer Guide.

### Ressourcenbasierte Berechtigungen

Wenn Sie AWS Glue in Verbindung mit Hive, Spark oder Presto in Amazon EMR verwenden, unterstützt AWS Glue ressourcenbasierte Richtlinien zur Steuerung des Zugriffs auf Datenkatalogressourcen. Zu diesen Ressourcen gehören Datenbanken, Tabellen, Verbindungen

und benutzerdefinierte Funktionen. Weitere Informationen finden Sie unter [Verwenden von ressourcenbasierten Richtlinien für AWS Glue](#) im AWS -Glue-Entwicklerhandbuch.

Wenn Sie ressourcenbasierte Richtlinien verwenden, um den Zugriff auf AWS Glue von Amazon EMR aus zu beschränken, muss der Principal, den Sie in der Berechtigungsrichtlinie angeben, der Rollen-ARN sein, der dem EC2-Instance-Profil zugeordnet ist, das bei der Erstellung eines Clusters angegeben wird. Beispielsweise können Sie für eine ressourcenbasierte Richtlinie, die an einen Katalog angehängt ist, den Rollen-ARN für die Standarddienstrolle für Cluster-EC2-Instances, `EMR_EC2_DefaultRole`, als angeben, indem Sie das im folgenden Beispiel gezeigte Principal Format verwenden:

```
arn:aws:iam::acct-id:role/EMR_EC2_DefaultRole
```

Die *Konto-ID* kann sich von der AWS Glue-Konto-ID unterscheiden. Dies ermöglicht den Zugriff von EMR-Clustern in verschiedenen Konten aus. Sie können mehrere Principals angeben, von denen jeder aus einem anderen Konto stammt.

## Überlegungen zur Verwendung von AWS Glue Data Catalog

Beachten Sie die folgenden Punkte, wenn Sie den AWS Glue-Datenkatalog als Metastore mit Hive verwenden:

- Das Hinzufügen von zusätzlichen JAR-Dateien mithilfe der Hive-Shell wird nicht unterstützt. Um dieses Problem zu umgehen, verwenden Sie die `hive-site`-Konfigurationsklassifizierung zum Festlegen der `hive.aux.jars.path`-Eigenschaft. Auf diese Weise werden zusätzliche JAR-Dateien in den Hive-Klassenpfad hinzugefügt.
- [Hive-Transaktionen](#) werden nicht unterstützt.
- Das Umbenennen von Tabellen in AWS Glue wird nicht unterstützt.
- Wenn Sie eine Hive-Tabelle erstellen, ohne eine `LOCATION` anzugeben, werden die Tabellendaten an dem in der Eigenschaft `hive.metastore.warehouse.dir` angegebenen Ort gespeichert. Standardmäßig ist dies ein Speicherort in HDFS. Wenn ein anderer Cluster auf die Tabelle zugreifen muss, schlägt er fehl, sofern er nicht über ausreichende Berechtigungen für den Cluster verfügt, der die Tabelle erstellt hat. Da HDFS-Speicher vorübergehend ist, gehen außerdem die Tabellendaten verloren, wenn der Cluster beendet wird, und die Tabelle muss neu erstellt werden. Wir empfehlen, dass Sie `LOCATION` in Amazon S3 angeben, wenn Sie eine Hive-Tabelle mit AWS Glue erstellen. Alternativ können Sie die `hive-site`-Konfigurationsklassifizierung verwenden, um einen Speicherort in Amazon S3 für `hive.metastore.warehouse.dir`

anzugeben, der für alle Hive-Tabellen gilt. Wenn eine Tabelle an einem HDFS-Speicherort erstellt wurde und der Cluster, der sie erstellt hat, noch läuft, können Sie den Tabellenspeicherort von AWS Glue aus auf Amazon S3 aktualisieren. Weitere Informationen finden Sie unter [Arbeiten mit Tabellen auf der AWS Glue-Konsole](#) im AWS Glue-Entwicklerhandbuch.

- Partitionswerte, die Anführungszeichen und Apostrophe enthalten, werden nicht unterstützt, zum Beispiel `PARTITION (owner="Doe 's")`.
- [Spaltenstatistiken](#) werden für emr-5.31.0 und höher unterstützt.
- Die Verwendung der [Hive-Autorisierung](#) wird nicht unterstützt. Als Alternative sollten Sie die Verwendung [ressourcenbasierter Richtlinien für AWS Glue](#) in Betracht ziehen. Weitere Informationen finden Sie unter [Verwenden von ressourcenbasierten Richtlinien für Amazon EMR Access to AWS Glue Data Catalog](#).
- [Hive-Einschränkungen](#) werden nicht unterstützt.
- Die [kostenbasierte Optimierung in Hive](#) wird nicht unterstützt.
- Die Einstellung `hive.metastore.partition.inherit.table.properties` wird nicht unterstützt.
- Die Verwendung der folgenden Metastore-Konstanten wird nicht unterstützt:  
BUCKET\_COUNT, BUCKET\_FIELD\_NAME, DDL\_TIME, FIELD\_TO\_DIMENSION,  
FILE\_INPUT\_FORMAT, FILE\_OUTPUT\_FORMAT, HIVE\_FILTER\_FIELD\_LAST\_ACCESS,  
HIVE\_FILTER\_FIELD\_OWNER, HIVE\_FILTER\_FIELD\_PARAMS, IS\_ARCHIVED,  
META\_TABLE\_COLUMNS, META\_TABLE\_COLUMN\_TYPES, META\_TABLE\_DB,  
META\_TABLE\_LOCATION, META\_TABLE\_NAME, META\_TABLE\_PARTITION\_COLUMNS,  
META\_TABLE\_SERDE, META\_TABLE\_STORAGE, ORIGINAL\_LOCATION.
- Wenn Sie einen Prädikatausdruck verwenden, müssen sich explizite Werte auf der rechten Seite des Vergleichsoperators befinden. Andernfalls schlagen Abfragen möglicherweise fehl.
  - Richtig: `SELECT * FROM mytable WHERE time > 11`
  - Falsch: `SELECT * FROM mytable WHERE 11 > time`
- Die Amazon-EMR-Versionen 5.32.0 und 6.3.0 und höher unterstützen die Verwendung benutzerdefinierter Funktionen (UDFs) in Prädikatausdrücken. Wenn Sie frühere Versionen verwenden, schlagen Ihre Abfragen möglicherweise fehl, weil Hive versucht, die Abfrageausführung zu optimieren.
- [Temporäre Tabellen](#) werden nicht unterstützt.
- Wir empfehlen, Tabellen mithilfe von Anwendungen über Amazon EMR zu erstellen, anstatt sie direkt mit AWS Glue zu erstellen. Das Erstellen einer Tabelle mit AWS Glue kann dazu führen, dass erforderliche Felder fehlen und Abfrageausnahmen auftreten.

- In EMR 5.20.0 oder höher wird das parallel Partitionsbereinigen automatisch für Spark und Hive aktiviert, wenn AWS Glue Data Catalog als Metastore verwendet wird. Diese Änderung reduziert die Zeit für die Abfrageplanung erheblich, da mehrere Anfragen parallel ausgeführt werden, um Partitionen abzurufen. Die Gesamtzahl der Segmente, die gleichzeitig ausgeführt werden können, liegt zwischen 1 und 10. Der Standardwert ist 5, was eine empfohlene Einstellung ist. Sie können dies ändern, indem Sie die Eigenschaft `aws.glue.partition.num.segments` in der `hive-site`-Konfigurationsklassifizierung angeben. Wenn eine Drosselung auftritt, können Sie das Feature ausschalten, indem Sie den Wert auf 1 ändern. Weitere Informationen finden Sie unter [AWS -Glue-Segmentstruktur](#).

## Verwenden einer externen MySQL-Datenbank oder von Amazon Aurora;

Damit Sie eine externe MySQL-Datenbank oder Amazon Aurora als Hive-Metastore verwenden können, überschreiben Sie die Standardkonfigurationswerte für den Metastore in Hive, um den Speicherort der externen Datenbank anzugeben, entweder auf einer Amazon-RDS-MySQL-Instance oder einer Amazon-Aurora-PostgreSQL-Instance.

### Note

Von Hive werden gleichzeitige Schreibzugriffe auf Metastore-Tabellen weder unterstützt noch verhindert. Wenn zwei Cluster gemeinsam Metastore-Daten nutzen, müssen Sie sicherstellen, dass Sie nicht gleichzeitig in dieselbe Metastore-Tabelle schreiben, es sei denn, sie schreiben auf verschiedene Partitionen derselben Metastore-Tabelle.

Das folgende Verfahren zeigt, wie Sie die Standardkonfigurationswerte für den Speicherort des Hive-Metastores überschreiben und einen Cluster mit dem neu konfigurierten Metastore-Speicherort starten.

So erstellen Sie einen Metastore außerhalb des EMR-Clusters

1. Erstellen Sie eine MySQL- oder Aurora-PostgreSQL-Datenbank. Wenn Sie PostgreSQL verwenden, müssen Sie es konfigurieren, nachdem Sie Ihren Cluster bereitgestellt haben. Bei der Clustererstellung wird nur MySQL unterstützt. Informationen zu den Unterschieden zwischen Aurora MySQL und Aurora PostgreSQL finden Sie unter [Überblick über Amazon Aurora MySQL](#) und [Arbeiten mit Amazon Aurora PostgreSQL](#). Weitere Informationen zum Erstellen einer Amazon-RDS-Datenbank finden Sie unter <https://aws.amazon.com/rds/>.

2. Ändern Sie Ihre Sicherheitsgruppen, um JDBC-Verbindungen zwischen Ihrer Datenbank und der -Master-Sicherheitsgruppe zuzulassen. ElasticMapReduce Informationen zum Ändern Ihrer Sicherheitsgruppen für den Zugriff finden Sie unter [Arbeiten mit von Amazon EMR verwalteten Sicherheitsgruppen](#).
3. Legen Sie die JDBC-Konfigurationswerte in `hive-site.xml` fest:

**⚠ Wichtig**

Wenn Sie sensible Daten wie z. B. Passwörter an die Amazon-EMR-Konfigurations-API übergeben, werden diese Informationen für Konten mit ausreichenden Berechtigungen angezeigt. Wenn Sie besorgt sind, dass diese Daten anderen Benutzern angezeigt werden könnten, erstellen Sie den Cluster mit einem administrativen Konto und schränken Sie den Zugriff anderer Benutzer (IAM-Benutzer oder Benutzer mit delegierten Anmeldeinformationen) auf die Services im Cluster ein, indem Sie eine Rolle erstellen, die Berechtigungen für den API-Schlüssel `elasticmapreduce:DescribeCluster` ausdrücklich verweigert.

- a. Erstellen Sie eine Konfigurationsdatei namens `hiveConfiguration.json`, die Änderungen an `hive-site.xml` enthält, wie im folgenden Beispiel gezeigt.

Ersetzen eines *Hostname* ist die DNS-Adresse der Amazon-RDS-Instance, auf der die Datenbank ausgeführt wird. *Benutzername* und *Password* sind die Anmeldeinformationen für Ihre Datenbank. Weitere Informationen zum Verbinden mit MySQL- und Aurora-Datenbank-Instances finden Sie unter [Verbinden mit einer DB-Instance, die die MySQL Database Engine ausführt](#) und [Verbinden mit einem DB-Cluster](#) im Amazon-RDS-Benutzerhandbuch.

`javax.jdo.option.ConnectionURL` ist die JDBC-Verbindungszeichenfolge für einen JDBC-Metastore. `javax.jdo.option.ConnectionDriverName` ist der Treiberklassenname für einen JDBC-Metastore.

Die MySQL-JDBC-Treiber werden von Amazon EMR installiert.

Die Werteigenschaft darf keine Leerzeichen oder Wagenrückläufe enthalten. Sie sollte in einer Zeile erscheinen.

```
[  
  {
```




```

    "Classification": "hive-site",
    "Properties": {
      "javax.jdo.option.ConnectionURL": "jdbc:mysql://hostname:3306/hive?
createDatabaseIfNotExist=true",
      "javax.jdo.option.ConnectionDriverName": "org.mariadb.jdbc.Driver",
      "javax.jdo.option.ConnectionUserName": "username",
      "javax.jdo.option.ConnectionPassword": "password"
    }
  }
]

```

- b. Verweisen Sie bei der Erstellung des Clusters auf die `hiveConfiguration.json` Datei, wie im folgenden Befehl gezeigt. AWS CLI In diesem Befehl ist die Datei lokal gespeichert. Sie können die Datei auch zu Amazon S3 hochladen und dort referenzieren, beispielsweise `s3://DOC-EXAMPLE-BUCKET/hiveConfiguration.json`.

 Note

Linux-Zeilenfortsetzungszeichen (`\`) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (`^`).

```

aws emr create-cluster --release-label emr-7.1.0 --instance-type m5.xlarge --
instance-count 2 \
--applications Name=Hive --configurations file://hiveConfiguration.json --use-
default-roles

```

4. Stellen Sie eine Verbindung mit dem Primärknoten Ihres Clusters her.

Informationen zum Verbinden mit dem Primärknoten finden Sie unter [Verbinden mit dem Primärknoten über SSH](#) im Verwaltungshandbuch für Amazon EMR.

5. Erstellen Sie Ihre Hive-Tabellen, indem Sie den Speicherort in Amazon S3 durch Eingabe eines Befehls ähnlich dem folgenden angeben:

```

CREATE EXTERNAL TABLE IF NOT EXISTS table_name
(
  key int,
  value int
)

```

```
LOCATION s3://DOC-EXAMPLE-BUCKET/hdfs/
```

6. Fügen Sie dem aktiven Cluster Ihr Hive-Skript hinzu.

Ihr Hive-Cluster wird unter Verwendung des Metastores in Amazon RDS ausgeführt. Starten Sie alle zusätzlichen Hive-Cluster, die diesen Metastore nutzen, indem Sie den Metastore-Speicherort angeben.

## Den Hive-JDBC-Treiber verwenden

Sie können beliebige Business-Intelligence-Tools wie Microsoft Excel, MicroStrategy, QlikView und Tableau mit Amazon EMR verwenden, um Daten zu erkunden und zu visualisieren. Viele dieser Tools erfordern einen Java Database Connectivity (JDBC)- oder einen Open Database Connectivity (ODBC)-Treiber. Amazon EMR unterstützt JDBC- und ODBC-Verbindungen.

Das folgende Beispiel zeigt die Verwendung von SQL Workbench/J als SQL-Client zum Verbinden eines Hive-Clusters in Amazon EMR. Weitere Treiber finden Sie unter [Verwenden von Business-Intelligence-Tools mit Amazon EMR](#).

Bevor Sie SQL Workbench/J installieren und damit arbeiten, laden Sie das Treiber-Paket herunter, und installieren Sie den Treiber. Die in dem Paket enthaltenen Treiber unterstützen die Hive-Versionen, die in den Amazon-EMR-Versionen 4.0 und höher verfügbar sind. Detaillierte Versionshinweise und Dokumentationen finden Sie in der PDF-Dokumentation in dem Paket.

- Das neueste Hive-JDBC-Treiberpaket herunterladen

<http://awssupportdatasvcs.com/bootstrap-actions/Simba/latest/>

- Ältere Versionen des Hive-JDBC-Treibers

<http://awssupportdatasvcs.com/bootstrap-actions/Simba/>

So installieren und konfigurieren Sie die SQL Workbench

1. Laden Sie über <http://www.sql-workbench.net/downloads.html> den SQL Workbench/J-Client für Ihr Betriebssystem herunter.
2. Installieren Sie SQL Workbench/J. Weitere Informationen finden Sie unter [Installieren und Starten von SQL Workbench/J](#) im Benutzerhandbuch für SQL Workbench/J.

3. Linux-, Unix-, Mac OS X-Benutzer: Erstellen Sie mithilfe des folgenden Befehls in einer Terminalsitzung einen SSH-Tunnel zum Master-Knoten Ihres Clusters. Ersetzen Sie *master-public-dns-name* durch den öffentlichen DNS-Namen des Hauptknotens und *path-to-key-file* mit dem Speicherort und dem Dateinamen Ihrer Datei für den privaten Amazon-EC2-Schlüssel (.pem).

```
ssh -o ServerAliveInterval=10 -i path-to-key-file -N -L 10000:localhost:10000  
hadoop@master-public-dns-name
```

Windows-Benutzer: Erstellen Sie in einer PuTTY-Sitzung einen SSH-Tunnel zum Hauptknoten Ihres Clusters (mit lokaler Port-Weiterleitung) und legen Sie 10000 für Quellport und *master-public-dns-name*:10000 für Ziel fest. Ersetzen Sie *master-public-dns-name* durch den öffentlichen DNS-Namen des Master-Knotens.

4. Fügen Sie SQL Workbench den JDBC-Treiber hinzu.
  - a. Klicken Sie im Dialogfeld Select Connection Profile (Verbindungsprofil auswählen) auf Manage Drivers (Treiber verwalten).
  - b. Klicken Sie auf das Symbol (leere Seite) Create a new entry (Neuen Eintrag erstellen).
  - c. Geben Sie in das Feld Name **Hive JDBC** ein.
  - d. Klicken Sie bei Library (Bibliothek) auf das Symbol Select the JAR file(s) (JAR-Datei(en) auswählen).
  - e. Navigieren Sie zu dem Speicherort, der die extrahierten Treiber enthält. Wählen Sie die Treiber aus, die in der heruntergeladenen Version des JDBC-Treiberpakets enthalten sind, und klicken Sie auf Öffnen.

Ihr JDBC-Treiberpaket kann beispielsweise die folgenden JAR-Dateien enthalten.

```
hive_metastore.jar  
hive_service.jar  
HiveJDBC41.jar  
libfb303-0.9.0.jar  
libthrift-0.9.0.jar  
log4j-1.2.14.jar  
ql.jar  
slf4j-api-1.5.11.jar  
slf4j-log4j12-1.5.11.jar  
TCLIServiceClient.jar  
zookeeper-3.4.6.jar
```

- f. Wählen Sie im Dialogfeld Please select one driver (Einen Treiber auswählen) die Optionen `com.amazon.hive.jdbc41.HS2Driver` und OK (OK) aus.
5. Wenn Sie zum Dialogfeld Manage Drivers (Treiber verwalten) zurückkehren, überprüfen Sie, ob das Feld Classname (Klassenname) ausgefüllt ist. Klicken Sie dann auf OK (OK).
6. Wenn Sie zum Dialogfeld Select Connection Profile (Verbindungsprofil auswählen) zurückkehren, überprüfen Sie, ob das Dialogfeld Driver (Treiber) auf Hive JDBC (Hive JDBC) festgelegt ist. Geben Sie die folgende JDBC-Verbindungszeichenfolge im Feld URL (URL) ein:  
`jdbc:hive2://localhost:10000/default`.
7. Wählen Sie OK (OK) aus, um die Verbindung herzustellen. Nachdem die Verbindung aufgebaut ist, werden oben im SQL Workbench/J-Fenster Verbindungsdetails angezeigt.

Weitere Informationen zum Verwenden von Hive und der JDBC-Schnittstelle finden Sie unter [HiveClient](#) und [HiveJDBCInterface](#) in der Apache Hive-Dokumentation.

## Verbessern der Hive-Leistung

Amazon EMR bietet Features, mit denen die Leistung optimiert werden kann, wenn Hive für Abfragen und Lese- bzw. Schreiboperationen über Daten in Amazon S3 verwendet wird.

S3 Select kann in einigen Anwendungen die Abfrageleistung bei CSV- und JSON-Dateien verbessern, indem die Verarbeitung an Amazon S3 ausgelagert wird.

Der EMRFS S3-optimierte Committer ist eine Alternative zu dieser [OutputCommitter](#)-Klasse, die Auflisten- und Umbenennungsvorgänge überflüssig macht, um die Leistung beim Schreiben von Dateien in Amazon S3 mit EMRFS zu verbessern.

Themen

- [Aktivierung des optimierten Hive EMRFS S3 Committers](#)
- [Verwendung von S3 Select mit Hive zur Verbesserung der Leistung](#)
- [MSCK-Optimierung](#)

## Aktivierung des optimierten Hive EMRFS S3 Committers

Der Hive EMRFS S3 Optimized Committer ist eine alternative Methode, mit der EMR Hive Dateien für Insert-Abfragen schreibt, wenn EMRFS verwendet wird. Der Committer macht Listen- und

Umbenennungsvorgänge überflüssig, die in Amazon S3 ausgeführt wurden, und verbessert die Leistung der Anwendung. Das Feature ist ab EMR 5.34 und EMR 6.5 verfügbar.

## Aktivieren des Committers

Wenn Sie die Verwendung von EMR Hive HiveEMRFSOptimizedCommitter zum Commit von Daten als Standard für alle von Hive verwalteten und externen Tabellen aktivieren möchten, verwenden Sie die folgende `hive-site`-Konfiguration in EM- 6.5.0- oder EMR-5.34.0-Clustern.

```
[
  {
    "classification": "hive-site",
    "properties": {
      "hive.blobstore.use.output-committer": "true"
    }
  }
]
```

### Note

Schalten Sie dieses Feature nicht ein, wenn `hive.exec.parallel` auf `true` eingestellt ist.

## Einschränkungen

Die folgenden grundlegenden Einschränkungen gelten für Tags (Markierungen):

- Die Aktivierung von Hive für das automatische Zusammenführen kleiner Dateien wird nicht unterstützt. Die standardmäßige Hive-Commit-Logik wird auch dann verwendet, wenn der optimierte Committer aktiviert ist.
- Hive-ACID-Tabellen werden nicht unterstützt. Die standardmäßige Hive-Commit-Logik wird auch dann verwendet, wenn der optimierte Committer aktiviert ist.
- Die Benennungsnomenklatur für geschriebene Dateien wurde von Hives von `<task_id>_<attempt_id>_<copy_n>` auf `<task_id>_<attempt_id>_<copy_n>_<query_id>` geändert. Zum Beispiel eine Datei mit dem Namen

`s3://warehouse/table/partition=1/000000_0` wird geändert zu `s3://warehouse/table/partition=1/000000_0-`

hadoop\_20210714130459\_ba7c23ec-5695-4947-9d98-8a40ef759222-1. Das `query_id` hier ist eine Kombination aus Benutzername, Zeitstempel und UUID.

- Wenn sich benutzerdefinierte Partitionen auf verschiedenen Dateisystemen (HDFS, S3) befinden, wird dieses Feature automatisch deaktiviert. Wenn sie aktiviert ist, wird die standardmäßige Hive-Commit-Logik verwendet.

## Verwendung von S3 Select mit Hive zur Verbesserung der Leistung

Mit Amazon-EMR-Version 5.18.0 und höher können Sie [S3 Select](#) mit Hive auf Amazon EMR verwenden. S3 Select ermöglicht es Anwendungen, nur eine Teilmenge von Daten aus einem Objekt abzurufen. Bei Amazon EMR wird die numerische Arbeit zur Filterung großer Datensätze von dem Cluster an Amazon S3 ausgelagert. Dies kann die Leistung in einigen Anwendungen verbessern und reduziert die Menge der zwischen Amazon EMR und Amazon S3 übertragenen Daten.

S3 Select wird mit Hive-Tabellen auf der Grundlage von CSV- und JSON-Dateien und durch die Einstellung der Konfigurationsvariablen `s3select.filter` auf `true` während der Hive-Sitzung unterstützt. Weitere Informationen und Beispiele finden Sie unter [Angeben von S3 Select in Ihrem Code](#).

### Ist S3 Select das Richtige für meine Anwendung?

Wir empfehlen, dass Sie Benchmark-Tests für Ihre Anwendungen im Vergleich mit und ohne S3 Select durchführen, um zu sehen, ob es für Ihre Anwendung geeignet sein könnte.

Verwenden Sie die folgenden Richtlinien, um zu bestimmen, ob Ihre Anwendung ein Kandidat für die Verwendung von S3 ist:

- Ihre Abfrage filtert mehr als die Hälfte des ursprünglichen Datensatzes.
- Ihre Abfrage-Filterprädikate verwenden Spalten mit einem Datentyp, der von Amazon S3 Select unterstützt wird. Weitere Informationen finden Sie unter [Datentypen](#) im Benutzerhandbuch für Amazon Simple Storage Service.
- Ihre Netzwerkverbindung zwischen Amazon S3 und dem Amazon-EMR-Cluster hat eine gute Übertragungsgeschwindigkeit und verfügbare Bandbreite. Amazon S3 komprimiert keine HTTP-Antworten, sodass die Antwortgröße bei komprimierten Eingabedateien wahrscheinlich zunimmt.

## Überlegungen und Einschränkungen

- Die Amazon-S3-serverseitige Verschlüsselung mit vom Kunden bereitgestellten Verschlüsselungsschlüsseln (SSE-C) und die clientseitige Verschlüsselung werden nicht unterstützt.
- Die Eigenschaft `AllowQuotedRecordDelimiters` wird nicht unterstützt. Wenn diese Eigenschaft angegeben ist, schlägt die Abfrage fehl.
- Nur CSV- und JSON-Dateien im UTF-8-Format werden unterstützt. Mehrzeilige CSVs und JSON werden nicht unterstützt.
- Nur unkomprimierte oder gzip- oder bzip2-Dateien werden unterstützt.
- Kommentarzeichen auf der letzten Zeile werden nicht unterstützt.
- Leere Zeilen am Ende einer Datei werden nicht verarbeitet.
- Hive auf Amazon EMR unterstützt die primitiven Datentypen, die S3 Select unterstützt. Weitere Informationen finden Sie unter [Datentypen](#) im Benutzerhandbuch für Amazon Simple Storage Service.

## Angeben von S3 Select in Ihrem Code

Um S3 Select in Ihrer Hive-Tabelle zu verwenden, erstellen Sie die Tabelle durch Angabe von `com.amazonaws.emr.s3select.hive.S3SelectableTextInputFormat` als `INPUTFORMAT`-Klassenname, und geben Sie einen Wert für die Eigenschaft `s3select.format` mit der `TBLPROPERTIES`-Klausel an.

S3 Select ist standardmäßig deaktiviert, wenn Sie Abfragen ausführen. Aktivieren Sie S3 Select, indem Sie `s3select.filter` auf `true` in Ihrer Hive-Sitzung setzen, wie unten gezeigt. Die folgenden Beispiele zeigen, wie Sie S3 Select angeben, wenn Sie eine Tabelle aus CSV- und JSON-Dateien erstellen und anschließend die Tabelle mit einer einfachen Select-Anweisung abfragen.

### Example CREATE-TABLE-Anweisung für CSV-basierte Tabelle

```
CREATE TABLE mys3selecttable (  
  col1 string,  
  col2 int,  
  col3 boolean  
)  
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','  
STORED AS  
INPUTFORMAT
```

```
'com.amazonaws.emr.s3select.hive.S3SelectableTextInputFormat'
OUTPUTFORMAT
'org.apache.hadoop.hive.q1.io.HiveIgnoreKeyTextOutputFormat'
LOCATION 's3://path/to/mycsvfile/'
TBLPROPERTIES (
  "s3select.format" = "csv",
  "s3select.headerInfo" = "ignore"
);
```

## Example CREATE-TABLE-Anweisung für JSON-basierte Tabelle

```
CREATE TABLE mys3selecttable (
  col1 string,
  col2 int,
  col3 boolean
)
ROW FORMAT SERDE 'org.apache.hive.hcatalog.data.JsonSerDe'
STORED AS
INPUTFORMAT
'com.amazonaws.emr.s3select.hive.S3SelectableTextInputFormat'
OUTPUTFORMAT
'org.apache.hadoop.hive.q1.io.HiveIgnoreKeyTextOutputFormat'
LOCATION 's3://path/to/json/'
TBLPROPERTIES (
  "s3select.format" = "json"
);
```

## Example SELECT-TABLE-Anweisung

```
SET s3select.filter=true;
SELECT * FROM mys3selecttable WHERE col2 > 10;
```

## MSCK-Optimierung

Hive speichert eine Liste von Partitionen für jede Tabelle in seinem Metastore. Wenn Partitionen jedoch direkt zum Dateisystem hinzugefügt oder daraus entfernt werden, bemerkt der Hive-Metastore diese Änderungen nicht. Der [MSCK-Befehl](#) aktualisiert die Partitionsmetadaten im Hive-Metastore für Partitionen, die direkt zum Dateisystem hinzugefügt oder daraus entfernt wurden. Die Syntax für den Befehl lautet:

```
MSCK [REPAIR] TABLE table_name [ADD/DROP/SYNC PARTITIONS];
```



Hive implementiert diesen Befehl wie folgt:

1. Hive ruft alle Partitionen für die Tabelle aus dem Metastore ab. Aus der Liste der Partitionspfade, die im Dateisystem nicht existieren, wird dann eine Liste von Partitionen erstellt, die aus dem Metastore gelöscht werden sollen.
2. Hive sammelt die im Dateisystem vorhandenen Partitionspfade, vergleicht sie mit der Liste der Partitionen aus dem Metastore und generiert eine Liste von Partitionen, die dem Metastore hinzugefügt werden müssen.
3. Hive aktualisiert den Metastore im Modus ADD, DROP oder SYNC.

#### Note

Wenn der Metastore viele Partitionen enthält, dauert die Ausführung des Schritts zur Überprüfung, ob eine Partition nicht im Dateisystem existiert, sehr lange, da der `exists-API`-Aufruf des Dateisystems für jede Partition erfolgen muss.

In Amazon EMR 6.5.0 führte Hive ein Flag mit dem Namen `hive.emr.optimize.msck.fs.check` ein. Wenn diese Option aktiviert ist, überprüft Hive anhand der Liste der Partitionspfade aus dem Dateisystem, die in Schritt 2 oben generiert wurde, ob eine Partition vorhanden ist, anstatt Dateisystem-API-Aufrufe durchzuführen. In Amazon EMR 6.8.0 hat Hive diese Optimierung standardmäßig aktiviert, sodass das Flag `hive.emr.optimize.msck.fs.check` nicht mehr gesetzt werden muss.

## Verwendung von Hive Live Long and Process (LLAP)

Amazon EMR 6.0.0 unterstützt die Live Long and Process (LLAP)-Funktionalität für Hive. LLAP verwendet persistente Daemons mit intelligentem In-Memory-Caching, um die Abfrageleistung im Vergleich zum vorherigen Standard-Tez-Container-Ausführungsmodus zu verbessern.

Die Hive LLAP-Daemons werden als YARN-Service verwaltet und ausgeführt. Da ein YARN-Service als lange laufende YARN-Anwendung betrachtet werden kann, sind einige Ihrer Cluster-Ressourcen für Hive LLAP dediziert und können nicht für andere Workloads verwendet werden. Weitere Informationen finden Sie unter [LLAP](#) und [YARN Service API \(YARN-Service-API\)](#).

## Hive LLAP auf Amazon EMR aktivieren

Wenn Sie Hive LLAP auf Amazon EMR aktivieren möchten, geben Sie beim Starten eines Clusters die folgende Konfiguration an.

```
[
  {
    "Classification": "hive",
    "Properties": {
      "hive.llap.enabled": "true"
    }
  }
]
```

Weitere Informationen finden Sie unter [Konfigurieren von Anwendungen](#).

Standardmäßig weist Amazon EMR etwa 60 Prozent der Cluster-YARN-Ressourcen Hive-LLAP-Daemons zu. Sie können den Anteil der dem Hive LLAP zugewiesenen Cluster-YARN-Ressource und die Anzahl der Aufgaben- und Core-Knoten, die für die Hive LLAP-Zuweisung berücksichtigt werden sollen, konfigurieren.

Die folgende Konfiguration startet beispielsweise Hive LLAP mit drei Daemons auf drei Aufgaben- oder Core-Knoten und weist 40 Prozent der YARN-Ressourcen der drei Kern- oder Aufgabenknoten den Hive LLAP-Daemons zu.

```
[
  {
    "Classification": "hive",
    "Properties": {
      "hive.llap.enabled": "true",
      "hive.llap.percent-allocation": "0.4",
      "hive.llap.num-instances": "3"
    }
  }
]
```

Sie können die folgenden `hive-site`-Konfigurationen in der Klassifikations-API verwenden, um Standard-LLAP-Ressoreneinstellungen außer Kraft zu setzen.

Property (Eigenschaft)	Description (Beschreibung)
hive.llap.daemon.yarn.container.mb	Gesamtgröße des LLAP-Daemons (in MB)
hive.llap.daemon.memory.per.instance.mb	Der Gesamtspeicher, der von Executoren im LLAP-Daemon-Container verwendet wird (in MB)
hive.llap.io.memory.size	Cachegröße für LLAP-Ein-/Ausgabe
hive.llap.daemon.num.executors	Anzahl der Executoren pro LLAP-Daemon

## Starten Sie LLAP manuell auf Ihrem Cluster

Alle von LLAP verwendeten Abhängigkeiten und Konfigurationen werden als Teil des Cluster-Startups in das LLAP-tar-Archiv gepackt. Wenn LLAP mithilfe von "hive.llap.enabled": "true" aktiviert wird, ist es empfehlenswert, Konfigurationsänderungen an LLAP mithilfe der Amazon-EMR-Neukonfiguration vorzunehmen.

Andernfalls müssen Sie bei manuellen Änderungen an `hive-site.xml`, das LLAP-tar-Archiv mithilfe des `hive --service llap`-Befehls neu erstellen, wie im folgenden Beispiel gezeigt wird.

```
# Define how many resources you want to allocate to Hive LLAP

LLAP_INSTANCES=<how many llap daemons to run on cluster>
LLAP_SIZE=<total container size per llap daemon>
LLAP_EXECUTORS=<number of executors per daemon>
LLAP_XMX=<Memory used by executors>
LLAP_CACHE=<Max cache size for IO allocator>

yarn app -enableFastLaunch

hive --service llap \
--instances $LLAP_INSTANCES \
--size ${LLAP_SIZE}m \
--executors $LLAP_EXECUTORS \
--xmx ${LLAP_XMX}m \
--cache ${LLAP_CACHE}m \
--name llap0 \
```

```
--auxhbase=false \  
--startImmediately
```

## Überprüfen Sie den LLAP-Status

Verwenden Sie den folgenden Befehl, um den Status von Hive LLAP über Hive zu überprüfen.

```
hive --service llapstatus
```

Verwenden Sie den folgenden Befehl, um den Status von Hive LLAP mit YARN zu überprüfen.

```
yarn app -status (name-of-llap-service)  
  
# example:  
yarn app -status llap0 | jq
```

## Starten oder stoppen Sie Hive LLAP

Da Hive LLAP als persistenter YARN-Service ausgeführt wird, müssen Sie den YARN-Service stoppen oder neu starten, um Hive LLAP zu stoppen oder neu zu starten. Die folgenden Befehle veranschaulichen dies.

```
yarn app -stop llap0  
yarn app -start llap0
```

## So ändern Sie die Größe der Hive LLAP-Daemons

Verwenden Sie den folgenden Befehl, um die Anzahl der LLAP-Instances zu reduzieren.

```
yarn app -flex llap0 -component llap -1
```

Weitere Informationen finden Sie unter [Flexible auf eine Komponente eines Services reagieren](#).

## Verschlüsselung im Hive

In diesem Abschnitt werden die Verschlüsselungstypen beschrieben, die Amazon EMR unterstützt.

## Modulare Parquet-Verschlüsselung in Hive

Die modulare Verschlüsselung von Parquet bietet Zugriffskontrolle und Verschlüsselung auf Spaltenebene, um den Datenschutz und die Datenintegrität für Daten zu verbessern, die im Parquet-Dateiformat gespeichert sind. Dieses Feature ist in Amazon EMR Hive ab Version 6.6.0 verfügbar.

Bisher unterstützte Lösungen für Sicherheit und Integrität, die das Verschlüsseln von Dateien oder das Verschlüsseln der Speicherebene beinhalten, werden im Verwaltungshandbuch für Amazon EMR unter [Verschlüsselungsoptionen](#) beschrieben. Diese Lösungen können für Parquet-Dateien verwendet werden, aber die Nutzung der neuen Features des integrierten Parquet-Verschlüsselungsmechanismus ermöglicht einen detaillierten Zugriff auf die Spaltenebene sowie Verbesserungen in Bezug auf Leistung und Sicherheit. Weitere Informationen zu diesem Feature finden Sie auf der Apache-Github-Seite [Parquet Modular Encryption](#).

Benutzer geben Konfigurationen mithilfe von Hadoop-Konfigurationen an Parquet-Leser und -Writer weiter. Die detaillierten Konfigurationen für Benutzer, um Reader und Writer so zu konfigurieren, dass sie die Verschlüsselung aktivieren und erweiterte Features umschalten, sind unter [PARQUET-1854: Eigenschaftsgesteuerte Schnittstelle zu Parquet Encryption Management](#) dokumentiert

### Verwendungsbeispiele

Das folgende Beispiel behandelt das Erstellen und Schreiben in eine Hive-Tabelle mit AWS KMS, die zur Verwaltung von Verschlüsselungsschlüsseln verwendet wird.

1. Implementieren Sie einen KmsClient für den AWS KMS Service, wie im Dokument [PARQUET-1373: Encryption Key Management Tools](#) beschrieben. Das folgende Beispiel zeigt einen Implementierungsausschnitt.

```
package org.apache.parquet.crypto.keytools;

import com.amazonaws.AmazonClientException;
import com.amazonaws.AmazonServiceException;
import com.amazonaws.regions.Regions;
import com.amazonaws.services.kms.AWSKMS;
import com.amazonaws.services.kms.AWSKMSClientBuilder;
import com.amazonaws.services.kms.model.DecryptRequest;
import com.amazonaws.services.kms.model.EncryptRequest;
import com.amazonaws.util.Base64;
import org.apache.hadoop.conf.Configuration;
import org.apache.parquet.crypto.KeyAccessDeniedException;
import org.apache.parquet.crypto.ParquetCryptoRuntimeException;
```

```
import org.apache.parquet.crypto.keytools.KmsClient;
import org.slf4j.Logger;
import org.slf4j.LoggerFactory;

import java.nio.ByteBuffer;
import java.nio.charset.Charset;
import java.nio.charset.StandardCharsets;

public class AwsKmsClient implements KmsClient {

    private static final AWSKMS AWSKMS_CLIENT = AWSKMSClientBuilder
        .standard()
        .withRegion(Regions.US_WEST_2)
        .build();
    public static final Logger LOG = LoggerFactory.getLogger(AwsKmsClient.class);

    private String kmsToken;
    private Configuration hadoopConfiguration;

    @Override
    public void initialize(Configuration configuration, String kmsInstanceID, String
kmsInstanceURL, String accessToken) throws KeyAccessDeniedException {
        hadoopConfiguration = configuration;
        kmsToken = accessToken;
    }

    @Override
    public String wrapKey(byte[] keyBytes, String masterKeyIdentifier) throws
KeyAccessDeniedException {
        String value = null;
        try {
            ByteBuffer plaintext = ByteBuffer.wrap(keyBytes);

            EncryptRequest req = new
EncryptRequest().withKeyId(masterKeyIdentifier).withPlaintext(plaintext);
            ByteBuffer ciphertext = AWSKMS_CLIENT.encrypt(req).getCiphertextBlob();

            byte[] base64EncodedValue = Base64.encode(ciphertext.array());
            value = new String(base64EncodedValue, Charset.forName("UTF-8"));
        } catch (AmazonClientException ae) {
            throw new KeyAccessDeniedException(ae.getMessage());
        }
        return value;
    }
}
```

```

    }

    @Override
    public byte[] unwrapKey(String wrappedKey, String masterKeyIdentifier) throws
    KeyAccessDeniedException {
        byte[] arr = null;
        try {
            ByteBuffer ciphertext =
            ByteBuffer.wrap(Base64.decode(wrappedKey.getBytes(StandardCharsets.UTF_8)));
            DecryptRequest request = new
            DecryptRequest().withKeyId(masterKeyIdentifier).withCiphertextBlob(ciphertext);
            ByteBuffer decipheredtext =
            AWSKMS_CLIENT.decrypt(request).getPlaintext();
            arr = new byte[decipheredtext.remaining()];
            decipheredtext.get(arr);
        } catch (AmazonClientException ae) {
            throw new KeyAccessDeniedException(ae.getMessage());
        }
        return arr;
    }
}

```

- Erstellen Sie Ihre AWS KMS Verschlüsselungsschlüssel für die Fußzeile sowie für die Spalten, auf die Ihre IAM-Rollen Zugriff haben, wie unter Schlüssel [erstellen](#) im Entwicklerhandbuch beschrieben. AWS Key Management Service Die Standard-IAM-Rolle ist EMR\_ECS\_Default.
- Fügen Sie in der Hive-Anwendung auf einem Amazon-EMR-Cluster den obigen Client mithilfe der ADD JAR-Anweisung hinzu, wie in der [Apache-Hive-Resources-Dokumentation](#) beschrieben. Nachfolgend finden Sie eine Beispielanweisung:

```
ADD JAR 's3://location-to-custom-jar';
```

Eine alternative Methode besteht darin, das JAR mithilfe einer Bootstrap-Aktion zur auxlib-Datei von Hive hinzuzufügen. Im Folgenden finden Sie eine Beispielzeile, die der Bootstrap-Aktion hinzugefügt werden soll:

```
aws s3 cp 's3://location-to-custom-jar' /usr/lib/hive/auxlib
```

- Legen Sie die folgenden Konfigurationen fest:

```
set
parquet.crypto.factory.class=org.apache.parquet.crypto.keytools.PropertiesDrivenCryptoFactory;
set
parquet.encryption.kms.client.class=org.apache.parquet.crypto.keytools.AwsKmsClient;
```

5. Erstellen Sie eine Hive-Tabelle im Parquet-Format, geben Sie die AWS KMS Schlüssel in SERDEPROPERTIES an und fügen Sie einige Daten ein:

```
CREATE TABLE my_table(name STRING, credit_card STRING)
ROW FORMAT SERDE 'org.apache.hadoop.hive ql.io.parquet.serde.ParquetHiveSerDe'
WITH SERDEPROPERTIES (
  'parquet.encryption.column.key'=<aws-kms-key-id-for-column-1>: credit_card',
  'parquet.encryption.footer.key'='<aws-kms-key-id-for-footer>')
STORED AS parquet
LOCATION "s3://<bucket>/<warehouse-location>/my_table";

INSERT INTO my_table SELECT
java_method ('org.apache.commons.lang.RandomStringUtils','randomAlphabetic',5) as
name,
java_method ('org.apache.commons.lang.RandomStringUtils','randomAlphabetic',10) as
credit_card
from (select 1) x lateral view posexplode(split(space(100),' ')) pe as i,x;

select * from my_table;
```

6. Stellen Sie sicher, dass Sie die Daten nicht lesen können, wenn Sie am selben Speicherort eine externe Tabelle ohne Zugriff auf AWS KMS Schlüssel erstellen (z. B. wenn der Zugriff auf die IAM-Rolle verweigert wurde).

```
CREATE EXTERNAL TABLE ext_table (name STRING, credit_card STRING)
ROW FORMAT SERDE 'org.apache.hadoop.hive ql.io.parquet.serde.ParquetHiveSerDe'
STORED AS parquet
LOCATION "s3://<bucket>/<warehouse-location>/my_table";

SELECT * FROM ext_table;
```

7. Die letzte Anweisung sollte die folgende Ausnahme auslösen:

```
Failed with exception
java.io.IOException:org.apache.parquet.crypto.KeyAccessDeniedException: Footer key:
access denied
```



## Verschlüsselung während der Übertragung in 2 HiveServer

Ab Amazon EMR Version 6.9.0 ist HiveServer 2 (HS2) als Teil der Sicherheitskonfiguration TLS/SSL-fähig. [Verschlüsselung während der Übertragung in 2 HiveServer](#) Dies wirkt sich darauf aus, wie Sie eine Verbindung zu HS2 herstellen, das auf einem Amazon-EMR-Cluster mit aktivierter Verschlüsselung während der Übertragung ausgeführt wird. Um eine Verbindung zu HS2 herzustellen, müssen Sie die Parameterwerte TRUSTSTORE\_PATH und TRUSTSTORE\_PASSWORD in der JDBC-URL ändern. Die folgende URL ist ein Beispiel für eine JDBC-Verbindung für HS2 mit den erforderlichen Parametern:

```
jdbc:hive2://HOST_NAME:10000/  
default;ssl=true;sslTrustStore=TRUSTSTORE_PATH;trustStorePassword=TRUSTSTORE_PASSWORD
```

Verwenden Sie die unten stehenden Anweisungen für die Verschlüsselung auf dem Cluster oder außerhalb von Cluster 2. HiveServer

### On-cluster HS2 access

Wenn Sie mit dem Beeline-Client auf HiveServer 2 zugreifen, nachdem Sie eine SSH-Verbindung zum Primärknoten hergestellt haben, verweisen Sie auf die Suche `/etc/hadoop/conf/ssl-server.xml` nach den Parameterwerten und mithilfe von Konfiguration `TRUSTSTORE_PATH` und `TRUSTSTORE_PASSWORD`. `ssl.server.truststore.location`  
`ssl.server.truststore.password`

Die folgenden Beispielbefehle können Ihnen beim Abrufen dieser Konfigurationen helfen:

```
TRUSTSTORE_PATH=$(sed -n '/ssl.server.truststore.location/,+2p' /etc/hadoop/conf/  
ssl-server.xml | awk -F "[><]" '/value/{print $3}')  
TRUSTSTORE_PASSWORD=$(sed -n '/ssl.server.truststore.password/,+2p' /etc/hadoop/  
conf/ssl-server.xml | awk -F "[><]" '/value/{print $3}')
```

### Off-cluster HS2 access

Wenn Sie von einem Client außerhalb des Amazon EMR-Clusters auf HiveServer 2 zugreifen, können Sie einen der folgenden Ansätze verwenden, um das und abzurufen: `TRUSTSTORE_PATH`  
`TRUSTSTORE_PASSWORD`

- Konvertieren Sie die PEM-Datei, die während der [Sicherheitskonfiguration erstellt wurde, in eine JKS-Datei und verwenden](#) Sie dieselbe in der JDBC-Verbindungs-URL. Verwenden Sie für openssl und keytool die folgenden Befehle:

```
openssl pkcs12 -export -in trustedCertificates.pem -inkey privateKey.pem -out
trustedCertificates.p12 -name "certificate"
keytool -importkeystore -srckeystore trustedCertificates.p12 -srcstoretype pkcs12
-destkeystore trustedCertificates.jks
```

- Alternativ können Sie auf `/etc/hadoop/conf/ssl-server.xml` verweisen, um die Parameterwerte `TRUSTSTORE_PATH` und `TRUSTSTORE_PASSWORD` mithilfe der Konfiguration `ssl.server.truststore.location` und `ssl.server.truststore.password` zu finden. Laden Sie die Truststore-Datei auf den Client-Computer herunter und verwenden Sie den Pfad auf dem Client-Computer als `TRUSTSTORE_PATH`.

Weitere Informationen zum Zugriff auf Anwendungen von einem Client außerhalb des Amazon-EMR-Clusters finden Sie unter [Verwenden des Hive-JDBC-Treibers](#).

## Versionsverlauf von Hive

In der folgenden Tabelle sind die Version von Hive aufgeführt, die in jeder Release-Version von Amazon EMR enthalten ist, zusammen mit den Komponenten, die mit der Anwendung installiert wurden. Informationen zu den Komponentenversionen in den einzelnen Versionen finden Sie im Abschnitt [Komponentenversion für Ihre Version](#) in [Amazon-EMR-7.x-Versionen](#), [Amazon-EMR-6.x-Versionen](#) oder [Amazon-EMR-5.x-Versionen](#).

Die Versionsinformationen von Hive

Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-5.36.2	2.3.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client,

Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
		hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn
emr-7.1.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, tez-on-worker, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-7.0.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, tez-on-worker, zookeeper-client, zookeeper-server
emr-6.15.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, tez-on-worker, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-6.14.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, tez-on-worker, zookeeper-client, zookeeper-server
emr-6.13.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, tez-on-worker, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-6.12.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, tez-on-worker, zookeeper-client, zookeeper-server
emr-6.11.1	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, tez-on-worker, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-6.11.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, tez-on-worker, zookeeper-client, zookeeper-server
emr-6.10.1	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, tez-on-worker, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-6.10.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, tez-on-worker, zookeeper-client, zookeeper-server
emr-6.9.1	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server



Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-6.9.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server
emr-6.8.1	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-6.8.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server
emr-6.7.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-5.36.1	2.3.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn
emr-5.36.0	2.3.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-6.6.0	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server
emr-5.35.0	2.3.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-6.5.0	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server
emr-6.4.0	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-6.3.1	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server
emr-6.3.0	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-6.2.1	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server
emr-6.2.0	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-6.1.1	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server
emr-6.1.0	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server



Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-6.0.1	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server
emr-6.0.0	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-5.34.0	2.3.8	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn
emr-5.33.1	2.3.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-5.33.0	2.3.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn
emr-5.32.1	2.3.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-5.32.0	2.3.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn
emr-5.31.1	2.3.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-5.31.0	2.3.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn
emr-5.30.2	2.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-5.30.1	2.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn
emr-5.30.0	2.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-5.29.0	2.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mysql-server, tez-on-yarn
emr-5.28.1	2.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mysql-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-5.28.0	2.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mysql-server, tez-on-yarn
emr-5.27.1	2.3.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn



Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-5.27.0	2.3.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.26.0	2.3.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-5.25.0	2.3.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.24.1	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-5.24.0	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.23.1	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-5.23.0	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.22.0	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-5.21.2	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.21.1	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-5.21.0	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.20.1	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-5.20.0	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.19.1	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-5.19.0	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.18.1	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn



Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-5.18.0	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.17.2	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-5.17.1	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.17.0	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-5.16.1	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.16.0	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-5.15.1	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.15.0	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-5.14.2	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.14.1	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-5.14.0	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.13.1	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-5.13.0	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.12.3	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-5.12.2	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.12.1	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn



Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-5.12.0	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.11.4	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-5.11.3	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.11.2	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-5.11.1	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.11.0	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-5.10.1	2.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.10.0	2.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-5.9.1	2.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.9.0	2.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-5.8.3	2.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.8.2	2.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-5.8.1	2.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.8.0	2.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-5.7.1	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.7.0	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn



Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-5.6.1	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.6.0	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-5.5.4	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.5.3	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-5.5.2	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.5.1	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-5.5.0	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.4.1	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-5.4.0	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.3.2	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-5.3.1	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn
emr-5.3.0	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-5.2.3	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn
emr-5.2.2	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-5.2.1	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn
emr-5.2.0	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn



Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-5.1.1	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn
emr-5.1.0	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-5.0.3	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn
emr-5.0.2	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-5.0.1	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn
emr-5.0.0	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-4.9.6	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server
emr-4.9.5	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server

Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-4.9.4	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server
emr-4.9.3	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server

Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-4.9.2	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource manager, hive-client, hcatalog-server, hive-server, mysql-server
emr-4.9.1	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource manager, hive-client, hcatalog-server, hive-server, mysql-server

Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-4.8.5	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server
emr-4.8.4	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server

Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-4.8.3	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server
emr-4.8.2	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server



Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-4.8.1	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server
emr-4.8.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server

Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-4.7.4	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server
emr-4.7.3	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server

Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-4.7.2	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server
emr-4.7.1	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server

Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-4.7.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server
emr-4.6.1	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, hive-server, mysql-server

Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-4.6.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httplibfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, hive-server, mysql-server
emr-4.5.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httplibfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, hive-server, mysql-server

Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-4.4.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httplibfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, hive-server, mysql-server
emr-4.3.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httplibfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, hive-server, mysql-server

Amazon-EMR-Version sbezeichnung	Hive-Version	Mit Hive installierte Komponenten
emr-4.2.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, hive-server, mysql-server
emr-4.1.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, hive-server, mysql-server
emr-4.0.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, hive-server, mysql-server

## Hive-Versionshinweise nach Version

### Themen

- [Amazon EMR 7.1.0 — Versionshinweise zu Hive](#)
- [Amazon EMR 6.15.0 – Versionshinweise zu Hive](#)
- [Amazon EMR 6.14.0 – Versionshinweise zu Hive](#)
- [Amazon EMR 6.13.0 – Versionshinweise zu Hive](#)
- [Amazon EMR 6.12.0 – Versionshinweise zu Hive](#)
- [Amazon EMR 6.11.0 – Versionshinweise zu Hive](#)
- [Amazon EMR 6.10.0 – Versionshinweise zu Hive](#)
- [Amazon EMR 6.9.0 – Versionshinweise zu Hive](#)
- [Amazon EMR 6.8.0 – Versionshinweise zu Hive](#)
- [Amazon EMR 6.7.0 – Versionshinweise zu Hive](#)
- [Amazon EMR 6.6.0 – Versionshinweise zu Hive](#)

### Amazon EMR 7.1.0 — Versionshinweise zu Hive

#### Amazon EMR 7.1.0 — Hive-Änderungen

Typ	Beschreibung
Fehlerbehebung	<a href="#">HIVE-24381</a> — Komprimierte Texteingabe gibt 0 Zeilen zurück, wenn die Kopf-/Fußzeile übersprungen wird.
Fehlerbehebung	<a href="#">HIVE-24190</a> — LLAP: könnte laut TEZ-4233 DISK_ERROR_EXCEPTION zurückgeben. ShuffleHandler
Fehlerbehebung	<a href="#">HIVE-23073</a> — Shade Netty.
Fehlerbehebung	<a href="#">HIVE-23073</a> — Shade Netty und Upgrade auf Netty 4.1.48.Final.



Typ	Beschreibung
Fehlerbehebung	<a href="#">HIVE-23148</a> — Der externe Llap-Clientfluss ist aufgrund von Netty Shading unterbrochen.
Fehlerbehebung	<a href="#">HIVE-25180</a> — Aktualisiert Netty.
Fehlerbehebung	<a href="#">HIVE-24524</a> — LLAP ShuffleHandler: Führen Sie ein Upgrade auf Netty4 durch und entfernen Sie die Netty3-Abhängigkeit von Hive, wo dies möglich ist.
Fehlerbehebung	<a href="#">HIVE-28000</a> — Hive QL: Die „not in“ - Klausel liefert falsche Ergebnisse, wenn kein Typenzwang stattfinden kann.
Fehlerbehebung	<a href="#">HIVE-27993</a> — Netty4 sollte einen Boss-Thread verwenden. ShuffleHandler
Upgrade	Führt ein Upgrade von Netty auf 4.1.100.Final durch
Upgrade	Führt ein Upgrade von Jetty auf 9.4.53.v2 0231009 durch
Upgrade	Führt ein Upgrade von Zookeeper auf 3.9.1 durch

## Amazon EMR 7.1.0 — Hive-Änderungen

- Amazon EMR 7.1 aktualisiert Hive auf Netty 4.1.100.Final, um die Sicherheitslücken in Netty3 zu beheben. Da Hive von Netty3 abhängig ist, hive-druid-handler hat Hive das hive-druid-handler JAR nicht im Klassenpfad von Hive in Amazon EMR 7.1. Eine bevorstehende Amazon EMR-Version wird es in den Klassenpfad von Hive aufnehmen, sobald der Druid-Handler 4.1.100.Final oder spätere Versionen von Netty unterstützt. Wenden Sie sich an den AWS Support, wenn Sie das hive-druid-handler JAR in Amazon EMR-Versionen 7.1 oder höher benötigen.

## Amazon EMR 6.15.0 – Versionshinweise zu Hive

### Amazon EMR 6.15.0 – Hive-Änderungen

Typ	Beschreibung
Funktion	Support für <a href="#">TEZ-4397</a> – Für das asynchrone Öffnen von Tez-Splits unterstützt Hive jetzt die unter <a href="#">Asynchrone Öffnung von Tez-Splits</a> beschriebenen Tez-Konfigurationen.
Fehlerbehebung	<a href="#">HIVE-25400</a> – Verschieben Sie die Offset-Aktualisierung in <code>BytesColumnVector</code> nach <code>setValPreallocated</code> .
Fehlerbehebung	<a href="#">HIVE-25190</a> – Korrigiert viele kleine Zuordnungen in <code>BytesColumnVector</code> .
Upgrade	Apache Hadoop auf 3.3.6 aktualisieren.
Upgrade	<a href="#">HIVE-26684</a> – Upgrade von <code>maven-shade-plugin</code> auf 3.4.1.
Verbesserung	Um die Startzeit des Amazon-EMR-Clusters zu reduzieren, entfernen Sie 15 Sekunden Ruhezeit aus dem <code>HCatalog-Startup-Skript</code> .

## Amazon EMR 6.14.0 – Versionshinweise zu Hive

### Amazon EMR 6.14.0 – Hive-Änderungen

Typ	Beschreibung
Verbesserung	<a href="#">HIVE-26762</a> : Entfernen des Kürzens von Operanden in <code>HiveFilterSetOptTransponseRule</code>

Typ	Beschreibung
Fehlerbehebung	<a href="#">HIVE-27582</a> : Zwischenspeichern Sie das Eingabeformat der HBase-Tabelle nicht in FetchOperator
Fehlerbehebung	<a href="#">HIVE-26452</a> : NPE bei der Konvertierung von JOIN in MAPJOIN und JOIN eine Spalte, auf die mehr als einmal verwiesen wird
Fehlerbehebung	<a href="#">HIVE-26416</a> : AcidUtils.isRawFormatFile() überträgt InvalidProtocolBufferException zur non-ORC-Datei
Fehlerbehebung	<a href="#">HIVE-26105</a> : Spalten anzeigen zeigt zusätzliche Werte an, wenn die Spalte Kommentare ein bestimmtes chinesisches Schriftzeichen enthält
Fehlerbehebung	<a href="#">HIVE-25864</a> : Die Hive-Abfrageoptimierung erzeugt einen falschen Plan für die Übertragung von Prädikaten mit Fensterfunktion
Fehlerbehebung	<a href="#">HIVE-25224</a> : Fehler bei mehreren INSERT-Anweisungen mit Tabellen mit unterschiedlichen bucketing_versions -Ergebnissen
Fehlerbehebung	<a href="#">HIVE-24151</a> : MultiDelimiterDe verschiebt Daten, wenn Zeichenfolgen non-ASCII-Zeichen enthalten
Fehlerbehebung	<a href="#">HIVE-23606</a> : (LLAP) Verzögerung bei DirectByteBuffer bei der Bereinigung für EncodedReaderImpl

Typ	Beschreibung
Fehlerbehebung	<a href="#">HIVE-22165</a> : Die durch <a href="#">HIVE-14296</a> <code>SessionManager.closeSession</code> eingeführte Synchronisation verursacht eine hohe Latenz auf einem ausgelasteten Hive-Server
Fehlerbehebung	<a href="#">HIVE-21304</a> : Machen Sie die Nutzung der Bucketing-Version robuster

## Amazon EMR 6.13.0 – Versionshinweise zu Hive

### Amazon EMR 6.13.0 – Hive-Änderungen

Typ	Beschreibung
Verbesserung	Aktualisieren Sie Python-Skripte zur Support von Python3
Verbesserung	<a href="#">HIVE-27097</a> : Verbessern Sie die Wiederholungsstrategie für Client und Server MetaStore
Fehlerbehebung	<a href="#">HIVE-21778</a> : CBO: „Strukt ist nicht Null“ wird als NULL-fähig bewertet, was immer zu einem Fehlschlagen des Filters in der Abfrage führt
Fehlerbehebung	<a href="#">HIVE-21009</a> : Dem Benutzer wird die Möglichkeit hinzugefügt, einen Bind-Benutzer festzulegen
Fehlerbehebung	<a href="#">HIVE-22661</a> : Die Komprimierung schlägt bei Tabellen ohne Bucket fehl, deren Daten im Pfad geladen wurden
Fehlerbehebung	<a href="#">HIVE-19718</a> : Beim Hinzufügen von Partitionen in großen Mengen wird auch die Tabelle für jede Partition abgerufen

Typ	Beschreibung
Fehlerbehebung	<a href="#">HIVE-22173</a> : Eine Abfrage mit mehreren seitlichen Ansichten hängt während der Kompilierung
Fehlerbehebung	<a href="#">HIVE-27088</a> : Falsche Ergebnisse, wenn innere und äußere Verknüpfungen mit Post-Join-Filtern zusammengeführt werden
Fehlerbehebung	<a href="#">HIVE-21935</a> : Hive-Vektorisierung: Eingeschränkte Leistung bei vektorisiertem UDF
Fehlerbehebung	<a href="#">HIVE-25299</a> : Das Umwandeln von Zeitstempeln in numerische Datentypen ist für Nicht-UTC-Zeitzone falsch
Fehlerbehebung	<a href="#">HIVE-24626</a> : LLAP: Leser-Threads könnten ausgehungert werden, wenn alle I/O-Elevator-Threads ausgelastet sind, um andere Leser mit voller Warteschlange in die Warteschlange zu stellen
Fehlerbehebung	<a href="#">HIVE-27029</a> : Die Hive-Abfrage schlägt mit dem Fehler Dateisystem geschlossen fehl. Die Überarbeitung für HIVE-26352 wurde abgeschlossen
Fehlerbehebung	<a href="#">HIVE-26352</a> : Die Prüfung des Zugriffs auf die Tez-Warteschlange schlägt mit einer GSS-Ausnahme bei der Komprimierung fehl
Fehlerbehebung	<a href="#">HIVE-24590</a> : Bei der Protokollierung von Vorgängen gibt es immer noch Log4j-Appender
Fehlerbehebung	<a href="#">HIVE-24552</a> : Mögliche Leckagen oder Anhäufung von HMS-Verbindungen in loadDynamicPartitions

Typ	Beschreibung
Fehlerbehebung	<a href="#">HIVE-27069</a> : Falsche Ergebnisse beim Bucket-Map-Join
Fehlerbehebung	<a href="#">HIVE-27344</a> : Fügt einen Nullcheck in #close hinzu RecordReaderImpl
Fehlerbehebung	<a href="#">HIVE-27439</a> : Support Leerzeichen in Dezimalzahlen
Fehlerbehebung	<a href="#">HIVE-27267</a> : Falsche Ergebnisse beim Bucket-Map-Join für eine Spalte mit Dezimalzahlen und Unterabfrage
Fehlerbehebung	<a href="#">HIVE-21986</a> : HiveServer Web-UI: Einstellung von Strict-Transport-Security im Standard-Antwort-Header
Fehlerbehebung	<a href="#">HIVE-22148</a> : S3A-Delegierungstoken werden der Job-Konfiguration des Compactors nicht hinzugefügt.
Fehlerbehebung	<a href="#">HIVE-22622</a> : Hive ermöglicht es, eine Struktur mit doppelten Attributnamen zu erstellen
Fehlerbehebung	<a href="#">HIVE-22008</a> : Der LIKE-Operator sollte einer mehrzeiligen Eingabe entsprechen
Fehlerbehebung	<a href="#">HIVE-23144</a> : LLAP: Lassen Sie die Bereinigung auf ServiceStop durchführen QueryTracker
Fehlerbehebung	<a href="#">HIVE-22391</a> : NPE beim Überprüfen des Zwischenspeichers für Hive-Abfrageergebnisse
Fehlerbehebung	<a href="#">HIVE-23305</a> : aufgrund von Rennbedingungen in AddNode NullPointerException LlapTaskSchedulerService

Typ	Beschreibung
Fehlerbehebung	<a href="#">HIVE-22178</a> : Parkett wirft hinterher FilterPredicate CastException SchemaEvolution
Fehlerbehebung	<a href="#">HIVE-21517</a> : Reparieren AggregateStatsCache
Fehlerbehebung	<a href="#">HIVE-21825</a> : Verbessern Sie die Client-Fehlermeldung, wenn Aktiv/Passive HA aktiviert ist
Fehlerbehebung	<a href="#">HIVE-23389</a> : kann dazu führen FilterMergeRule AssertionError
Fehlerbehebung	<a href="#">HIVE-22767</a> : Beeline analysiert Semikolons in Kommentaren nicht richtig
Fehlerbehebung	<a href="#">HIVE-22996</a> : Beim BasicStats Parsen sollte proaktiv nach Null oder einer leeren Zeichenfolge gesucht werden
Fehlerbehebung	<a href="#">HIVE-22808</a> : <u>behandelt nicht</u> HiveRelFieldTrimmer HiveTableFunctionScan
Fehlerbehebung	<a href="#">HIVE-22437</a> : LLAP-Metadaten-Cache NPE beim Sperren von Metadaten.
Fehlerbehebung	<a href="#">HIVE-22606</a> : protokolliert avro.schema.literal auf der INFO-Ebene AvroSerde
Fehlerbehebung	<a href="#">HIVE-22713</a> : Für die Join-Fil (*) -RS-Struktur sollte keine konstante Weitergabe durchgeführt werden
Fehlerbehebung	<a href="#">HIVE-21624</a> : LLAP: CPU-Metriken auf Thread-Ebene sind defekt
Fehlerbehebung	<a href="#">HIVE-22815</a> : Reduziert die unnötige Erstellung von Dateisystemobjekten in mOutput

Typ	Beschreibung
Fehlerbehebung	<a href="#">HIVE-23060</a> : Die Abfrage schlägt mit dem Fehler fehl: „Der Ausdruck für Gruppierungssätze ist nicht im GROUP BY-Schlüssel enthalten. In der Nähe des Tokens ist ein Fehler aufgetreten
Fehlerbehebung	<a href="#">HIVE-22236</a> : Ansicht konnte nicht erstellt werden, wenn Ansicht ausgewählt wird, die die NOT IN-Unterabfrage enthält
Fehlerbehebung	<a href="#">HIVE-19886</a> : Protokolle können an zwei Dateien weitergeleitet werden, wenn – hiveconf hive.log.file verwendet wird
Fehlerbehebung	<a href="#">HIVE-20620</a> : Kollisionen beim Einfügen in sortierte MM-Tabellen mit dynamischer Partitionierung manifestieren
Fehlerbehebung	<a href="#">HIVE-14557</a> : Nullpointer, wenn sowohl als auch Mapjoin aktiviert SkewJoin sind
Fehlerbehebung	<a href="#">HIVE-20471</a> : Probleme beim Abrufen des Standard-Datenbankpfads
Fehlerbehebung	<a href="#">HIVE-20598</a> : Tippfehler HiveAlgorithmsUtil in Berechnungen korrigieren
Fehlerbehebung	<a href="#">HIVE-14737</a> : Problem beim Zugriff auf /logs in einer Kerberized Hive Server 2-Weboberfläche
Fehlerbehebung	<a href="#">HIVE-20733</a> : GenericUdfopEqualNS darf = in Planbeschreibungen nicht verwenden
Fehlerbehebung	<a href="#">HIVE-20848</a> : Nach dem Einstellen schlägt die Abfrage mit der Meldung „Tabelle nicht gefunden“ fehl. UpdateInputAccessTimeHook



Typ	Beschreibung
Fehlerbehebung	<a href="#">HIVE-18929</a> : Die Methode <code>humanReadableInt</code> in <code>HiveStringUtils.java</code> hat eine Race-Bedingung.
Fehlerbehebung	<a href="#">HIVE-20841</a> : LLAP: Machen Sie dynamische Ports konfigurierbar
Fehlerbehebung	<a href="#">HIVE-20930</a> : <code>VectorCoalesce</code> Wird im FILTER-Modus nicht wirksam
Fehlerbehebung	<a href="#">HIVE-21007</a> : Semi Join + Union kann zu falschen Plänen führen
Fehlerbehebung	<a href="#">HIVE-21074</a> : Das Bereinigen von Abfragen in Hive-Bucket-Tabellen funktioniert nicht, wenn die Bedingung <code>IS NOT NULL</code> vorliegt
Fehlerbehebung	<a href="#">HIVE-21223</a> : <code>CachedStore</code> gibt eine Null-Partition zurück, wenn die Partition nicht existiert
Fehlerbehebung	<a href="#">HIVE-19625</a> : Potenzieller NPE und Ausblenden der tatsächlichen Ausnahme in Hive <code>#copyFiles</code>
Fehlerbehebung	<a href="#">HIVE-17020</a> : Aggressive RS-Deduplizierung kann fälschlicherweise den Zweig des OP-Baums entfernen
Fehlerbehebung	HIVE-20168: <a href="#">Protokollierung versteckt</a> ReduceSinkOperator
Fehlerbehebung	<a href="#">HIVE-20879</a> : Die Verwendung von Null in einem Projektionsausdruck führt zu <code>CastException</code>
Fehlerbehebung	<a href="#">HIVE-20888</a> TxnHandler: <code>sort ()</code> wurde für unveränderliche Listen aufgerufen

Typ	Beschreibung
Fehlerbehebung	<a href="#">HIVE-19948</a> : teilt den Befehl nicht richtig durch HiveCli Semikolon auf, wenn sich in der Zeichenfolge Anführungszeichen befinden
Fehlerbehebung	<a href="#">HIVE-20621</a> : Wird in resultset.next aufgerufen, was zu inkrementeller Langsamkeit führt GetOperationStatus
Fehlerbehebung	<a href="#">HIVE-20854</a> : Sensible Standardeinstellungen: Das Zookeeper-Heartbeat-Intervall von Hive beträgt 20 Minuten, ändern Sie es auf 2
Fehlerbehebung	<a href="#">HIVE-20330</a> : H kann nicht mehrere Objekte für einen Job mit mehreren Eingaben verarbeiten CatLoader InputJobInfo
Fehlerbehebung	<a href="#">HIVE-20787</a> : Der DummyRow-Fall behandelt die Wiederverwendung nicht MapJoinBy tesTableContainer
Fehlerbehebung	<a href="#">HIVE-20331</a> : Die Abfrage mit Union All, Lateral View und Join schlägt mit der Fehlermeldung „Kann kein übergeordnetes Element im untergeordneten Operator gefunden“ fehl.
Fehlerbehebung	<a href="#">HIVE-19968</a> : Die UDF-Ausnahme wird nicht ausgelöst
Fehlerbehebung	<a href="#">HIVE-20410</a> : Abgebrochenes Einfügen und Überschreiben in der Transaktionstabelle verursacht den Fehler „Es ist nicht genug Verlauf verfügbar für ...“
Fehlerbehebung	<a href="#">HIVE-20059</a> : Hive-Streaming sollte in Ausnahmefällen bedingungslos das Shade-Präfix ausprobieren

Typ	Beschreibung
Fehlerbehebung	<a href="#">HIVE-19424</a> : NPE-Eingang MetaDataFormatters
Fehlerbehebung	<a href="#">HIVE-20355</a> : Den Parameter von.setSchema bereinigen HiveConnection
Fehlerbehebung	<a href="#">HIVE-20858</a> : Der Serializer wurde mit der Konfiguration in Utilities nicht korrekt initialisiert. createEmptyBuckets
Fehlerbehebung	<a href="#">HIVE-20424</a> : Das Schematool darf den Beeline-Verlauf nicht verunreinigen
Fehlerbehebung	<a href="#">HIVE-20338</a> : LLAP: Erzwingt synthetische Datei-IDs für Dateisysteme, die HDFS-Protokoll-Impls mit POSIX-Mutationssemantik haben
Fehlerbehebung	<a href="#">HIVE-11708</a> : Logische Operatoren werden mit NULL ausgelöst ClassCastExceptions
Fehlerbehebung	<a href="#">HIVE-21082</a> : In HPL/SQL unterstützt die Declare-Anweisung keine Variablen vom Typ Character
Fehlerbehebung	<a href="#">HIVE-16690</a> : Konfigurieren Sie das kartesische Tez-Produkt Edge auf der Grundlage der LLAP-Clustergröße
Fehlerbehebung	<a href="#">HIVE-21296</a> : Beim Löschen der Varchar-Partition wird eine Ausnahme ausgelöst
Fehlerbehebung	<a href="#">HIVE-14516</a> : OrcInputFormat SplitGenerator. Intern anrufen
Fehlerbehebung	<a href="#">HIVE-20981</a> : Streaming/Lecks AbstractRecordWriter HeapMemoryMonitor

Typ	Beschreibung
Fehlerbehebung	<a href="#">HIVE-20043</a> : 2: HiveServer hat einen statischen Synchronisierungsblock um einen SessionState AtomicBoolean
Fehlerbehebung	<a href="#">HIVE-20191</a> : Die PreCommit Patch-Anwendung schlägt nicht fehl, wenn der Patch leer ist
Fehlerbehebung	<a href="#">HIVE-20400</a> : Tabelle erstellen sollte immer einen vollständig qualifizierten Pfad verwenden , um mögliche FS-Mehrdeutigkeiten zu vermeiden
Fehlerbehebung	Fügen Sie eine Nullprüfung für SkewedInfo hinzu, bevor Sie auf schiefe Spalten zugreifen

## Amazon EMR 6.12.0 – Versionshinweise zu Hive

### Amazon EMR 6.12.0 – Hive-Änderungen

Typ	Beschreibung
Verbesserung	Support für JDK-11- und JDK-17-Laufzeit hinzugefügt
Verbesserung	Unterstützung für die Abfrage von Spaltennamen mit Berücksichtigung von Groß- und Kleinschreibung und reservierten Schlüsselwörtern hinzugefügt, wenn S3 Select verwendet wird. Um es zu verwenden, definieren Sie die Tabelleneigenschaft im Format „s3select .column.mapping“ = " <i>column1:fieldName1</i> , <i>column2:fieldName2</i> , ..."

Typ	Beschreibung
Verbesserung	<a href="#">HIVE-23133</a> : Numerische Operationen können je nach Hardware-Architektur zu unterschiedlichen Ergebnissen führen
Verbesserung	<a href="#">HIVE-27145</a> : Wird StrictMath für die übrigen mathematischen Funktionen als Nachfolger von HIVE-23133 verwendet
Fehlerbehebung	Behebung der Wildcard-Inkompatibilität in den HMS-APIs <code>get_partitions_by_filter</code> und <code>get_num_partitions_by_filter</code> , die durch die Portierung von <a href="#">HIVE-22900</a> in EMR Hive 6.4.0 verursacht wurde
Fehlerbehebung	<a href="#">HIVE-26736</a> : Autorisierungsfehler für verschachtelte Views mit WITH-Klausel
Fehlerbehebung	<a href="#">HIVE-22416</a> : MR-bezogene Betriebsprotokolle fehlen, wenn die parallel Ausführung aktiviert ist
Fehlerbehebung	<a href="#">HIVE-19653</a> : Falscher Prädikat-Pushdown für Groupby mit Gruppierungssätzen
Fehlerbehebung	HIVE-22094: Abfragen schlagen fehl mit: <a href="#">hive.ql.exec.vector</a> . <code>ClassCastException</code> <code>DecimalColumnVector</code> kann nicht in <code>Hive.QL.Exec.Vector.Decimal64</code> umgewandelt werden <code>ColumnVector</code>
Fehlerbehebung	<a href="#">HIVE-26340</a> : Der vektorisierte PTF-Operator schlägt fehl, wenn die Abfrage eine Fensterfunktion in Großbuchstaben hat
Fehlerbehebung	<a href="#">HIVE-26184</a> : <code>COLLECT_SET</code> mit <code>GROUP BY</code> ist sehr langsam, wenn einige Tasten stark schief sind

Typ	Beschreibung
Fehlerbehebung	<a href="#">ClassCastException HIVE-26373: beim Lesen von Zeitstempeln aus der HBase-Tabelle mit Avro-Daten</a>
Fehlerbehebung	<a href="#">HIVE-26388</a> : ClassCastException wenn die Quelltable der CTAS-Abfrage eine Spalte enthält, die keine Zeichenfolge ist, Upgrade <a href="#">HIVE-26172: Hive — Upgrade von Ant auf 1.10.11 aufgrund von CVE-2021-36373 und CVE-2021-36374</a>
Fehlerbehebung	<a href="#">HIVE-26114</a> : Wenn Sie die JDBC-Verbindung hiveserver2 mit dem Befehl dfs mit Präfixraum reparieren, wird dies zu einer Ausnahme führen
Fehlerbehebung	<a href="#">HIVE-26396</a> : Die Trunc-Funktion hat ein Problem mit dem Präzisionsabfangen und das Ergebnis hat viele 0
Fehlerbehebung	<a href="#">TablesWritten HIVE-26446</a> : HiveProto LoggingHook Das Feld für partitionierte Tabellen kann nicht gefüllt werden.
Fehlerbehebung	<a href="#">HIVE-26639</a> : und sollte sich nicht auf den Standardzeichensatz verlassen ConstantVectorExpression ExplainTask
Fehlerbehebung	<a href="#">HIVE-22670</a> : ArrayIndexOutOfBoundsException wenn der vektorisierte Reader zum Lesen einer Parquet-Datei verwendet wird
Fehlerbehebung	<a href="#">HIVE-23607</a> : Berechtigungsproblem: Das Erstellen einer Ansicht in einer anderen Ansicht ist erfolgreich, aber das Ändern der Ansicht schlägt fehl

Typ	Beschreibung
Fehlerbehebung	<a href="#">HIVE-25498</a> : Eine Abfrage mit mehr als 31 verschiedenen Funktionen gibt ein falsches Ergebnis zurück
Fehlerbehebung	<a href="#">HIVE-25780</a> : erstellt mehr als 64 Gruppierungssätze II DistinctExpansion
Fehlerbehebung	<a href="#">HIVE-23868</a> : Spezifikation der Fensterfunktion: unterstützt 0 vorne/nach
Fehlerbehebung	<a href="#">HIVE-24539</a> : Die Schemagenerierung sollte das Spaltentrennzeichen berücksichtigen OrclInputFormat
Fehlerbehebung	<a href="#">HIVE-23476</a> : LLAP: Ordnet Arenen auch für die Groß-/Kleinschreibung vorab zu
Fehlerbehebung	<a href="#">HIVE-25806</a> : Mögliches Leck in — Parquet, LLAP IO LlapCacheAwareFs
Fehlerbehebung	<a href="#">HIVE-23498</a> : Deaktivieren Sie die HTTP-Trace-Methode auf ThriftHttpClientService
Fehlerbehebung	<a href="#">HIVE-25729</a> : sollte benachrichtigt werden, wenn es vollständig initiiert ist ThriftUnionObjectInspector
Fehlerbehebung	<a href="#">HIVE-23846</a> : Vermeiden Sie unnötige Serialisierung und Deserialisierung von Bitvektoren
Fehlerbehebung	<a href="#">HIVE-24233</a> : außer dass die Unterabfrage bei deaktiviertem CBO einen Nullzeiger auslöst
Fehlerbehebung	<a href="#">HIVE-24276</a> : Sicherheitslücke in loggerconf.jsp Cross-Site Scripting HiveServer (XSS)

Typ	Beschreibung
Fehlerbehebung	<a href="#">HIVE-25721</a> : Das äußere Join-Ergebnis ist falsch
Fehlerbehebung	<a href="#">HIVE-25223</a> : Auswahl mit Limit gibt keine Zeilen in einer nicht systemeigenen Tabelle zurück
Fehlerbehebung	<a href="#">HIVE-25794</a> :: Anweisungen in einer Schleife protokollieren führt zu Speicherdruck CombineHiveRecordReader
Fehlerbehebung	<a href="#">HIVE-23602</a> : Verwenden Sie das Java Concurrent Package für Operation Handle Set
Fehlerbehebung	<a href="#">HIVE-24045</a> : Keine Protokollierung im Zusammenhang mit dem Zeitpunkt, zu dem die Standarddatenbank erstellt wird
Fehlerbehebung	<a href="#">HIVE-24305</a> : Das Avro-Dezimalschema füllt Skala/Genauigkeit nicht richtig aus, wenn der Wert in Anführungszeichen eingeschlossen ist
Fehlerbehebung	<a href="#">HIVE-25844</a> : Deserialisierungsfehler in Ausnahmefällen können dazu führen, dass Beeline sofort beendet wird
Fehlerbehebung	<a href="#">HIVE-25040</a> : Die Drop-Datenbank-Kaskade kann persistente Funktionen nicht entfernen
Fehlerbehebung	<a href="#">HIVE-23501</a> : AOOB, wenn komplexe Typen in primitive Typen umgewandelt werden VectorDeserializeRow
Fehlerbehebung	<a href="#">HIVE-23704</a> : Der Thrift-HTTP-Server verarbeitet et das Auth-Handle nicht korrekt



Typ	Beschreibung
Fehlerbehebung	<a href="#">HIVE-23529</a> : CTAS ist für Uniontype kaputt, wenn row_deserialize
Fehlerbehebung	<a href="#">HIVE-24144</a> : Die Zeichenfolge in gibt einen falschen Wert zurück getIdentifierQuote HiveDatabaseMetaData
Fehlerbehebung	<a href="#">HIVE-23850</a> : Erlaube PPD, wenn der Betreff keine Spalte mit vorhandenen Gruppierungssätzen ist
Fehlerbehebung	<a href="#">HIVE-24036</a> : Kryo-Ausnahme beim Serialisieren des Plans für den GetSplits-UDF-Aufruf
Fehlerbehebung	<a href="#">HIVE-25919</a> : ClassCastException beim Drücken eines booleschen Spaltenprädikats in HBaseStorageHandler
Fehlerbehebung	<a href="#">HIVE-25261</a> : Beim erneuten Versuch sollte HMSHandler das mit einer kurzen Beschreibung des Ziels umschließen MetaException
Fehlerbehebung	<a href="#">HIVE-24792</a> : Potenzielles Thread-Leck während des Betriebs
Fehlerbehebung	<a href="#">HIVE-23409</a> : Wenn das erneute Öffnen der TezSession Anwendung fehlschlägt, weil der Timeline-Dienst ausgefallen ist, wird das Standardformular nach einem erneuten Versuch geschlossen TezSession SessionPool
Fehlerbehebung	<a href="#">HIVE-23615</a> : Verweisen Sie nicht auf Nullzeiger in der Beeline-Befehlsklasse
Fehlerbehebung	<a href="#">HIVE-24849</a> : Erzeugt ein Socket-Timeout für externe Tabellen, wenn der Speicherort eine große Anzahl von Dateien enthält (betrifft 3.1.2)

Typ	Beschreibung
Fehlerbehebung	<a href="#">HIVE-24193</a> : Die Auswahlabfrage für die umbenannte Hive-Acid-Tabelle erzeugt keine Ausgabe
Fehlerbehebung	<a href="#">HIVE-25209</a> : SELECT-Abfrage mit SUM-Funktion, die zu einem unerwarteten Ergebnis führt
Fehlerbehebung	<a href="#">HIVE-23666</a> : checkHashMode Effizienz wird übersprungen, wenn für einen Gruppierungs-Operator keine Gruppierung festgelegt ist
Fehlerbehebung	<a href="#">HIVE-23873</a> : Die Abfrage der Hive JDBC-Tabelle schlägt mit NPE fehl, wenn CBO ausgeschaltet ist StorageHandler
Fehlerbehebung	<a href="#">HiveStreamingConnection HIVE-24149</a> : Schließt die HMS-Verbindung nicht
Fehlerbehebung	<a href="#">HIVE-25561</a> : Die abgebrochene Aufgabe sollte die Datei nicht übertragen. (betrifft die Versionen 2.x und 3.x)
Fehlerbehebung	<a href="#">HIVE-25683</a> : Schließt das Lesegerät. AcidUtils isRawFormatDatei
Fehlerbehebung	<a href="#">HIVE-24294</a> : TezSessionPool Sessions können ausgelöst werden AssertionError
Fehlerbehebung	<a href="#">HIVE-24182</a> : Ranger-Autorisierungsproblem mit permanenten UDFs
Fehlerbehebung	<a href="#">HIVE-22805</a> : Die Vektorisierung mit bedingtem Array oder Map ist nicht implementiert und löst einen Fehler aus

Typ	Beschreibung
Fehlerbehebung	<a href="#">HIVE-22828</a> : Decimal64: NVL- und CASE-Anweisungen konvertieren implizit decimal64 in 128
Fehlerbehebung	<a href="#">HIVE-21398</a> : Spalten mit geschätzten Statistiken sollten nicht als eindeutige Schlüssel betrachtet werden
Fehlerbehebung	<a href="#">HIVE-22490</a> : Das Hinzufügen von Jars mit Sonderzeichen in ihrem Pfad führt zu einem Fehler
Fehlerbehebung	<a href="#">HIVE-22700</a> : Bei nicht autorisierten Komprimierungen kann Speicherplatz verloren gehen
Fehlerbehebung	<a href="#">HIVE-22053</a> : Der Funktionsname wird beim Erstellen der Funktion nicht normalisiert
Fehlerbehebung	<a href="#">HIVE-22595</a> : Dynamische Partitionseinfügungen schlagen in einer Avro-Tabellentabelle mit externem Schema fehl
Fehlerbehebung	<a href="#">HIVE-21795</a> : Die Zeile mit der Rollup-Zusammenfassung fehlt möglicherweise, wenn ein Mapjoin in einer partitionierten Tabelle stattfindet
Fehlerbehebung	<a href="#">HIVE-22987</a> : in wann ist Null ClassCast Exception VectorCoalesce DataTypePhysicalVariation
Fehlerbehebung	<a href="#">HIVE-22814</a> : bei der Vektorisierung ArrayIndexOutOfBounds getData Type PhysicalVariation
Fehlerbehebung	<a href="#">HIVE-22523</a> : Der Fehlerhandler in blockiert möglicherweise, wenn die Warteschlange voll ist LlapRecordReader

Typ	Beschreibung
Fehlerbehebung	<a href="#">HIVE-21796</a> : ArrayWritableObjectInspector.equals kann O (2^nesting_depth) Zeit in Anspruch nehmen
Fehlerbehebung	<a href="#">HIVE-22929</a> : Leistung: Beim Parsen von Bezeichnern in Anführungszeichen wird Regex als Wegwerf-Regex über String.replaceAll() verwendet
Fehlerbehebung	<a href="#">HIVE-21641</a> : Der externe Llap-Client gibt Dezimalspalten mit unterschiedlicher Genauigkeit/Skalierung zurück als Beeline
Fehlerbehebung	<a href="#">HIVE-22207</a> : Tez: wird ausgelöst, wenn „dfs.block.size“ auf dem Cluster „128m“ ist SplitGenerator NumberFormatException
Fehlerbehebung	<a href="#">HIVE-22114</a> : Das Einfügen einer Abfrage für eine partitionierte Einfügetabelle schlägt fehl, wenn alle Buckets leer sind
Fehlerbehebung	<a href="#">HIVE-22165</a> : <a href="#">Die durch HIVE-14296</a> auf .closeSession eingeführte Synchronisation verursacht eine hohe Latenz auf einem ausgelasteten Hive-Server SessionManager
Fehlerbehebung	<a href="#">HIVE-22744</a> : Für den Scheitelpunkt mit mehr als einer Außenkante sollte ein proportionaler Sortierspeicher vorhanden sein TezTask
Fehlerbehebung	<a href="#">HIVE-22072</a> : Wenn Sie eine Tabelle ändern, um eine Spalte zu ändern, werden die Beschränkungsreferenzen nicht aktualisiert
Fehlerbehebung	<a href="#">HIVE-22075</a> : Korrigiert die max-reducers=1-Regression von HIVE-14200

Typ	Beschreibung
Fehlerbehebung	<a href="#">HIVE-22527</a> : Hive on Tez: Der Job zum Zusammenführen kleiner Dateien wird in eine andere Warteschlange gestellt (Standard-Warteschlange)
Fehlerbehebung	<a href="#">HIVE-22816</a> : QueryCache: Abfragen, die Views verwenden, können nach der CTE-Erweiterung zwischengespeichert werden
Fehlerbehebung	<a href="#">HIVE-22733</a> : Nach dem Deaktivieren der Betriebsprotokolleigenschaft in Hive speichert HS2 immer noch das Betriebsprotokoll
Fehlerbehebung	<a href="#">HIVE-22699</a> : Masken-UDFs sollten den numerischen Wert 0 maskieren
Fehlerbehebung	<a href="#">HIVE-23356</a> : Die Hash-Aggregation ist bei der Verarbeitung von Abfragen mit Gruppierungssatzausdrücken immer deaktiviert.
Fehlerbehebung	<a href="#">HIVE-21568</a> HiveRelOptUtil: isRowFilteringDer Plan sollte das Projekt überspringen
Fehlerbehebung	<a href="#">HIVE-21760</a> : Die Optimierung geteilter Arbeit sollte bei SMB-Joins umgangen werden
Fehlerbehebung	<a href="#">HIVE-22712</a> : Der ReExec Treiber führt die Abfrage unabhängig von der benutzerdefinierten Warteschlange aus und sendet die Abfrage in der Standardwarteschlange
Fehlerbehebung	<a href="#">HIVE-21397</a> : BloomFilter Für Hive Managed [ACID] funktioniert die Tabelle nicht wie erwartet

Typ	Beschreibung
Fehlerbehebung	<a href="#">HIVE-23011</a> : Der Optimierer für geteiltes Arbeiten sollte beim Vergleich von Joins die restlichen Prädikate überprüfen
Fehlerbehebung	<a href="#">HIVE-21412</a> : funktioniert nicht mit ACID-Tabellen PostExecOrcFileDump
Fehlerbehebung	<a href="#">HIVE-22201</a> : ConvertJoinMapJoin # wird ausgelöst, wenn keine große Tabelle checkShuffleSize ForLargeTable ausgewählt ArrayIndexOutOfBoundsException ist
Fehlerbehebung	<a href="#">HIVE-21971</a> : HS2 gibt den Classloader aufgrund von `:CONSTRUCTOR_CACHE` mit temporären Funktionen + GenericUdf preis ReflectionUtils
Fehlerbehebung	<a href="#">HIVE-21368</a> : Vektorisierung: HiveDecimal Unnötige Decimal64 -> Konvertierung
Fehlerbehebung	<a href="#">HIVE-25416: Hive-Metastore-Speicherleck aufgrund eines Fehlers</a> datanucleus-api-jdo
Fehlerbehebung	<a href="#">HIVE-22219</a> : Das Herunterfahren eines Node-Managers blockiert den Neustart des LLAP-Dienstes
Fehlerbehebung	<a href="#">HIVE-21793</a> : CBO ruft Spaltenstatistiken ab, auch wenn hive.stats.fetch.column.stats auf false gesetzt ist
Fehlerbehebung	<a href="#">HIVE-22163</a> : CBO: Durch die Aktivierung von CBO wird die Statistikschatzung aktiviert, auch wenn die Schätzung deaktiviert ist
Fehlerbehebung	<a href="#">HIVE-18735</a> : Beim Erstellen einer Tabelle wird das Transaktionsattribut verloren

Typ	Beschreibung
Fehlerbehebung	<a href="#">HIVE-22433</a> : Hive JDBC Storage Handler: Falsche Ergebnisse wurden von BOOLEAN und TIMESTAMP aus der JDBC-Datenquelle abgerufen DataType
Fehlerbehebung	<a href="#">ObjectStoreHIVE-19430</a> : cleanNotificationEvents OutOfMemory über eine große Anzahl ausstehender Ereignisse
Fehlerbehebung	<a href="#">HIVE-20785</a> : Falscher Schlüsselname im JDBC. DatabaseMetaData getPrimaryKeys Methode
Fehlerbehebung	<a href="#">HIVE-16116</a> : Beeline wirft NPE, wenn beeline.hiveconfvariables= {} in beeline.properties
Fehlerbehebung	<a href="#">HIVE-20066</a> : hive.load.data.owner wird mit dem vollständigen Prinzipal verglichen
Fehlerbehebung	<a href="#">HIVE-20489</a> : Erläutern Sie, dass der Abfrageplan nicht mehr funktioniert
Fehlerbehebung	<a href="#">HIVE-21033</a> : Wenn vergessen wird, den Vorgang zu beenden, werden weitere 2 Ausgänge unterbrochen HiveServer
Fehlerbehebung	<a href="#">HIVE-19888</a> : Irreführende Warnung „METASTORE_FILTER_HOOK wird ignoriert“ von SessionState
Fehlerbehebung	<a href="#">HIVE-20303</a> : ÜBERSCHREIBUNGSTABELLE EINFÜGEN db.TABLE PARTITION (...) WENN NICHT EXISTIERT, wirft InvalidTableException
Fehlerbehebung	<a href="#">HIVE-16144</a> : CompactionInfo hat kein Equals/ hashCode, wird aber in Set verwendet

Typ	Beschreibung
Fehlerbehebung	<a href="#">HIVE-20818</a> : Mit einer WHERE-Unterabfrage erstellte Ansichten betrachten Ansichten, auf die in der Unterabfrage verwiesen wird, als direkte Eingabe
Fehlerbehebung	<a href="#">HIVE-21005</a> : LLAP: Lesen von mehr Stripes pro Split-Leaks ZlibCodecs
Fehlerbehebung	<a href="#">HIVE-20771</a> : schlägt bei leeren Strukturen fehl. LazyBinarySerDe
Fehlerbehebung	<a href="#">HIVE-18852</a> : Irreführende Fehlermeldung bei der Validierung von Änderungen an Tabellen
Fehlerbehebung	<a href="#">HIVE-21124</a> : HPL/SQL unterstützt die Anweisung CREATE TABLE LIKE nicht
Fehlerbehebung	<a href="#">HIVE-20935</a> : Das Hochladen des LLAP-Paket-Tarballs schlägt in EC2 fehl, was zu einem Startfehler des LLAP-Dienstes führt
Fehlerbehebung	<a href="#">HIVE-20409</a> : Hive ACID: Durch Aktualisieren/Löschen/Zusammenführen wird das HDFS-Staging-Verzeichnis nicht bereinigt
Fehlerbehebung	<a href="#">HIVE-20570</a> : Die Vereinigung ALL mit hive.optimize.union.remove=true hat einen falschen Plan
Fehlerbehebung	<a href="#">HIVE-20421</a> : Unzulässige Zeicheneinheit '\ b' in hive-default.xml.template
Fehlerbehebung	<a href="#">HIVE-19133</a> : Die phasenweisen Leistungskennzahlen der HS2-WebUI werden nicht korrekt angezeigt



Typ	Beschreibung
Fehlerbehebung	<a href="#">HIVE-18977</a> : Das Auflisten von Partitionen führt bei JDO und Direct SQL zu unterschiedlichen Ergebnissen
Fehlerbehebung	<a href="#">HIVE-20034</a> : Aus Gründen der Abwärtskompatibilität werden Änderungen bei der Ausnahmebehandlung rückgängig gemacht MetaStore
Fehlerbehebung	<a href="#">HIVE-20672</a> : Bei der Anmeldung sollte jedes festgelegte Intervall gemeldet werden LlapTaskSchedulerService
Fehlerbehebung	<a href="#">HIVE-12812</a> : Aktivieren Sie <code>mapred.input.dir.recursive</code> standardmäßig, um die Vereinigung mit Aggregatfunktion zu unterstützen
Fehlerbehebung	<a href="#">HIVE-20147</a> : Hive-Streaming-Ingest begnügt sich mit synchronisierter Protokollierung
Fehlerbehebung	<a href="#">HIVE-19203</a> : Thread-Sicherheitsproblem in HiveMetaStore
Fehlerbehebung	<a href="#">HIVE-20091</a> : Tez: Fügen Sie Sicherheitssanmeldedaten für die Ausgabe hinzu FileSinkOperator
Fehlerbehebung	<a href="#">HIVE-16906</a> : Hive ATShook sollte vor der Verbindung zu ATS nach <code>yarn.timeline-service.enabled</code> suchen
Fehlerbehebung	<a href="#">HIVE-20714</a> : SHOW tblproperties für eine einzelne Eigenschaft gibt den Wert in der Namensspalte zurück

Typ	Beschreibung
Fehlerbehebung	<a href="#">HIVE-24730</a> : Shims-Klassen überschreiben Werte aus hive-site.xml und tez-site.xml im Hintergrund
Fehlerbehebung	<a href="#">HIVE-22055</a> : Die gewählte Anzahl liefert nach dem Laden von Daten aus einer Textdatei ein falsches Ergebnis

## Amazon EMR 6.11.0 – Versionshinweise zu Hive

### Amazon EMR 6.11.0 – Hive-Änderungen

Typ	Beschreibung
Verbesserung	Es wurde Unterstützung für das Löschen von Partitionen mit mehreren Threads hinzugefügt, um die Leistung beim Löschen von Partitionen zu verbessern
Verbesserung	Support das Lesen von codierten Hive-Abfragedateien
Verbesserung	Der Tez Shuffle Handler ist standardmäßig für Hive-on-Tez-Jobs aktiviert
Fehler	Es wurde eine Option hinzugefügt, um die deterministische Verteilung von Schlüsseln an Reducer zu aktivieren, um ein falsches Ergebnis zu korrigieren, wenn hive.groupby.skewindata aktiviert ist (berichtet in <a href="#">HIVE-20220</a> )
Fehler	Fehler bei der Statistikberechnung behoben, wenn der Standardpartitionsname konfiguriert war

Typ	Beschreibung
Fehler	Beachten Sie alle benutzerdefinierten SSL-Klassifizierungsparameter, die übergeben werden, wenn SSL standardmäßig für HiveServer 2 in einem Cluster mit aktivierter Verschlüsselung während der Übertragung konfiguriert ist
Backport	<a href="#">HIVE-23617</a> : Probleme mit der Speicher-API wurden behoben FindBug
Backport	<a href="#">HIVE-26408</a> : Vektorisierung: Die Deallokation von Scratch-Spalten wurde korrigiert, untergeordnete Spalten werden nicht als Ausgabe wiederverwendet ConstantVectorExpression
Backport	<a href="#">HiveConfig HIVE-23614</a> : Immer weiterleiten an removeTempOr DuplicateFiles
Backport	<a href="#">HIVE-23354</a> : Entfernen Sie die Überprüfung der Dateigrößensicherheit von compareTempOr DuplicateFiles
Backport	<a href="#">HIVE-20344</a> : Beim Werfen durch SBA wurde ein Fehler behoben. PrivilegeSynchronizer AccessControlException Außerdem wurde die Eigenschaft hive.privilege.synchronizer eingeführt, um den Privilege Synchronizer zu deaktivieren
Backport	<a href="#">HIVE-15826</a> : Support die Konfiguration von 'serialization.encoding' für alle SerDes
Backport	<a href="#">HIVE-18284</a> : Behebung von NPE beim Einfügen von Daten mit der Klausel 'distribute by' mit Dynpart-Sortieroptimierung

Typ	Beschreibung
Backport	<a href="#">HIVE-24930</a> : Der Kurzschluss Operator. setDone() von der untergeordneten Operation wird im vektorisierten Codepfad nicht verwendet (wenn childSize == 1)
Backport	<a href="#">HIVE-24523</a> : Der vektorisierte Lesepfad für berücksichtigt nicht die SERDEPROPERTIES für den Zeitstempel LazySimpleSerde
Backport	<a href="#">HIVE-23265</a> : Doppelte Rowsets werden zurückgegeben, wenn Limit und Offset gesetzt sind
Backport	<a href="#">HIVE-21492</a> : Die mit dem Thrift/Custom-Tool generierte Parquetdatei kann nicht gelesen werden VectorizedParquetRecordReader
Backport	<a href="#">HIVE-22540</a> : Vektorisierung: Decimal64-Spalten funktionieren nicht mit. Vectorize dBatchUtil makeLikeColumnVektor ()
Backport	<a href="#">HIVE-22588</a> : Leert die verbleibenden Zeilen für die restlichen Gruppierungssätze, wenn Sie den Gruppierungsmodus des Vektors wechseln
Backport	<a href="#">HIVE-22551</a> : BytesColumnVector InitBuffer sollte Vektor und Länge konsistent bereinigen
Backport	<a href="#">HIVE-22448</a> : CBO: Erweitert die Mehrfachzählung mit einem Gruppierungsschlüssel
Backport	<a href="#">HIVE-22248</a> : Behebt anhaltende Probleme mit Statistiken
Backport	<a href="#">HIVE-22210</a> : Bei der Vektorisierung können beim Filtern verwendete Berechnungsausgabe spalten wiederverwendet werden

Typ	Beschreibung
Backport	<a href="#">HIVE-21531</a> : Vektorisierung: Alle NULL-Hash codes werden nicht mit Murmur3 berechnet
Backport	<a href="#">HIVE-20419</a> : Vektorisierung: Verhindert die Mutation von nach der Verwendung in einem Hashmap-Schlüssel VectorPartitionDesc
Backport	<a href="#">ClassCastException HIVE-19388</a> VectorMap JoinCommonOperator : während der Initialisierung
Backport	<a href="#">HIVE-21584</a> : Java 11-Vorbereitung: Der Systemklassenlader ist keine URL ClassLoader
Backport	<a href="#">HIVE-25107</a> : Die Klassenpfadprotokollierung sollte auf DEBUG-Ebene erfolgen (#2271)
Backport	<a href="#">HIVE-22097</a> : Inkompatibles java.util. ArrayList für Java 11
Backport	<a href="#">HIVE-23938</a> : LLAP: JDK11 – einige JVM-Argumente, die sich auf die Rotation von GC-Logdateien beziehen, können nicht mehr verwendet werden
Backport	<a href="#">HIVE-26226</a> : Schließt jdk.tools dep aus Hive-Metastore in Upgrade-Acid aus
Backport	<a href="#">HIVE-17879</a> : Aktualisieren Sie das Datanucleus Maven-Plugin
Backport	<a href="#">HIVE-27004</a> : DateTimeFormatterBuilder # appendZoneText kann 'UTC+' in Java-Versionen höher als 8 nicht analysieren
Backport	<a href="#">HIVE-16812</a> : filtert keine Löschereignisse VectorizedOrcAcidRowBatchReader

Typ	Beschreibung
Backport	<a href="#">HIVE-17917</a> : VectorizedOrcAcidRowBatchReader computeOffsetAndBucket-Optimierung
Backport	<a href="#">HIVE-19985</a> : ACID: Überspringen Sie die Dekodierung der ROW_ID-Abschnitte für schreibgeschützte Abfragen
Backport	<a href="#">HIVE-20635</a> : filtert VectorizedOrcAcidRowBatchReader keine Löschergebnisse für Originaldateien
Upgrade	Upgrade auf Javadoc 3.3.1
Upgrade	Aktualisieren Sie Javassist auf 3.24.1-GA
Upgrade	Update auf 2.0.0-M14 apache-directory-server

## Neue Konfigurationen

Name	Klassifizierung	Beschreibung
hive.metastore.fs.drop.partition.threads	hive-site	Anzahl der Core-Threads im Threadpool der Drop-Partition.
hive.metastore.fs.drop.partition.keepalive.time	hive-site	Zeit in Sekunden, in der ein asynchroner Drop-Partition-Thread (aus dem Thread-Pool) im Leerlauf auf das Eintreffen einer neuen Aufgabe wartet, bevor er beendet wird.
hive.metastore.fs.drop.partition.threadpool.max.queue.size	hive-site	Maximale Warteschlangenlänge, die im Thread-Pool zum Löschen von Partitionen aus dem Dateisystem verwendet werden soll.

Name	Klassifizierung	Beschreibung
hive.groupby.enable.deterministic.distribution	hive-site	Ermöglichen Sie die deterministische Verteilung von Schlüsseln an Reducer. Beim Aufrufen der Funktion Rand, die für die zufällige Partitionierung verwendet wird, wird ein konstanter Startwert übergeben.
hive.privilege.synchronizer	hive-site	Ob die Rechte eines externen Autorisierers in 2 regelmäßig synchronisiert werden sollen. HiveServer
hive.cli.query.file.encoding	hive-site	Dateikodierung für alle Arten von Abfragedateien (Abfragedatei, Init-Abfragedatei, RC-Datei usw.), die in den CLI-Argumenten bereitgestellt werden.
hive.emr.tez.shuffle.enabled	hive-site	Hive-on-Tez-Aufträge verwenden jetzt standardmäßig tez_shuffle anstelle von mapreduce_shuffle als Standard-Shuffle-Handler.

## Veraltete Konfigurationen

Die folgenden Konfigurationseigenschaften sind aufgrund von [HIVE-23354](#) veraltet und werden in den Amazon-EMR-Versionen 6.11.0 und höher nicht mehr unterstützt.

Name	Standardwert
<code>hive.mapred.reduce.tasks.speculative.execution</code>	false
<code>tez.am.speculation.enabled</code>	false

## Amazon EMR 6.10.0 – Versionshinweise zu Hive

### Amazon EMR 6.10.0 – Hive-Änderungen

Typ	Beschreibung
Funktion	Aktivieren Sie die AWS Lake Formation basierte Zugriffskontrolle für Apache Hive-Abfragen (Schreiben) <a href="#">über IAM Passthrough</a> (HiveCLI/Steps API).
Verbesserung	Deaktivieren Sie die Konfiguration <code>hive.log.explain.output</code> standardmäßig, um die Protokollgröße zu reduzieren
Backport	<a href="#">HIVE-26408</a> : Vektorisierung: Die Deallokation von Scratch-Spalten wurde korrigiert, untergeordnete Spalten dürfen nicht als Ausgabe wiederverwendet werden ConstantVectorExpression
Backport	<a href="#">HIVE-22269</a> : Korrektur der falschen Anzahl von Reduktoren bei Einfügeabfragen mit dynamischer Partition aufgrund fehlender Statistiken, die durch <a href="#">HIVE-20703</a> verursacht wurden.
Backport	<a href="#">HIVE-22891</a> : Extrahieren im Nicht-LLAP-Ausführungsmodus überspringen PartitionDesc CombineHiveRecord



Typ	Beschreibung
Backport	<a href="#">HIVE-23804</a> : Eine Standarddatenbank für spaltenstatistikspezifische Tabellen im Hive-Metastore-Schema hinzufügen, um sie abwärtskompatibel zu machen
Backport	<a href="#">HIVE-25277</a> : Langsames Löschen von Hive-Partitionen für Cloud-Objektspeicher mit teuren ListFiles
Backport	<a href="#">HIVE-19202</a> : CBO ist aufgrund von in ausgefallenen. NullPointerException HiveAggregate isBucketedInput()
Backport	<a href="#">HIVE-19048</a> : Fix Beeline Initscript-Fehler werden ignoriert
Backport	<a href="#">HIVE-21085</a> : Die Registrierung für materialisierte Ansicht startet eine nicht externe TEZ-Sitzung
Backport	<a href="#">HIVE-21675</a> : CREATE VIEW IF NOT EXISTS gibt statt „OK“ einen Fehler zurück, wenn die Ansicht bereits existiert. Dies ist eine Regression von Hive 2.
Backport	<a href="#">HIVE-21646</a> : Tez: Verhindert das Umgehen des TezTasks Thread-Log-Kontextes
Backport	<a href="#">HIVE-22054</a> : Vermeiden Sie eine rekursive Auflistung, um zu überprüfen, ob ein Verzeichnis leer ist
Backport	<a href="#">HIVE-16587</a> : NPE beim Einfügen komplexer Typen mit verschachtelten Nullwerten
Backport	<a href="#">HIVE-22647</a> : Aktiviert standardmäßig den Sitzungspool

Typ	Beschreibung
Backport	<a href="#">HIVE-13288</a> : Verwirrende Ausnahmemeldung in <code>.localizeResource DagUtils</code>
Backport	<a href="#">HIVE-23870</a> : Optimieren Sie mehrere Textkonvertierungen in <code>WritableHiveCharObjectInspector getPrimitiveJavaObjekt/HiveCharWritable</code>
Backport	<a href="#">HIVE-21498</a> : Aktualisieren Sie Thrift auf 0.13.0
Backport	<a href="#">HIVE-24378</a> : Führende und nachfolgende Leerzeichen werden vor der Dezimalkonvertierung nicht entfernt
Backport	<a href="#">HIVE-21341</a> : Sinnvolle Standardwerte: <code>hive.server2.idle.operation.timeout</code> und <code>hive.server2.idle.session.timeout</code> sind zu hoch
Backport	<a href="#">HIVE-22465</a> : Fügen Sie <code>ssl conf</code> hinzu <code>TezConfigurationFactory</code>
Backport	<a href="#">HIVE-24710</a> : Optimieren Sie die PTF-Iteration für <code>count (*)</code> , um die CPU- und I/O-Kosten zu senken
Backport	<a href="#">HIVE-15406</a> : Erwägen Sie die Vektorisierung der neuen Funktion „ <code>Trunc</code> “
Backport	<a href="#">HIVE-21541</a> : Korrigiert fehlende ASF-Header von HIVE-15406
Backport	<a href="#">HIVE-24808</a> : Analysierte Daten zwischenspeichern
Backport	<a href="#">HIVE-24746</a> : PTF: kann bei der Reichweitenberechnung optimiert werden <code>Timestamp ValueBoundaryScanner</code>

Typ	Beschreibung
Backport	<a href="#">HIVE-25059</a> : Das Alter-Ereignis wird während der Replikation in ein Umbenennungereignis umgewandelt
Backport	<a href="#">HIVE-25142</a> : Beim erneuten Hash in der Map-Join-Fast-Hash-Tabelle werden große Schlüssel beschädigt
Backport	<a href="#">HIVE-23756</a> : Der package.jdo-Datei wurden weitere Einschränkungen hinzugefügt
Backport	<a href="#">HIVE-25150</a> : Tabulatorzeichen werden vor der Dezimalkonvertierung nicht entfernt, ähnlich wie Leerzeichen, das in HIVE-24378 behoben wurde
Backport	<a href="#">HIVE-25093</a> : date_format () UDF gibt die Ausgabe nur in der UTC-Zeitzone zurück
Backport	<a href="#">HIVE-25268</a> : date_format udf gibt falsche Ergebnisse für Daten vor 1900 zurück, wenn die lokale Zeitzone nicht UTC ist
Backport	<a href="#">HIVE-25338</a> : AIOBE in Conv UDF, wenn die Eingabe leer ist
Backport	<a href="#">HIVE-22400</a> : UDF-Minute mit Zeit gibt NULL zurück
Backport	<a href="#">HIVE-25058</a> : PTF: kann bei der Reichweiteberechnung optimiert werden pt2 - TimestampValueBoundaryScanner isDistanceGreater

Typ	Beschreibung
Backport	<a href="#">HIVE-25449</a> : datediff() gibt eine falsche Ausgabe aus, wenn es in einer TEZ-Aufgabe mit einer anderen Zeitzone als UTC ausgeführt wird
Backport	<a href="#">HIVE-23688</a> : Vektorisierung: Für eine Spalte vom Typ Map, die einen Nullwert enthält IndexArrayOutOfBoundsException
Backport	<a href="#">HIVE-22247</a> : HiveH wird ausgelöst, wenn die Aufgabenausgabe der Partition leer ist FileOutputFormat FileNotFoundException
Backport	<a href="#">HIVE-25570</a> : Hive sollte den vollständigen URL-Pfad zur Autorisierung für den Befehl insert overwrite location senden
Backport	<a href="#">HIVE-22903</a> : Die vektorisierte row_number() setzt die Zeilennummer nach einem Batch zurück, falls ein konstanter Ausdruck in der Partitionsklausel vorliegt
Backport	<a href="#">HIVE-25549</a> : Falsche Ergebnisse für Fensterfunktionen mit Ausdruck in der PARTITION-BY- oder ORDER-BY-Klausel
Backport	<a href="#">HIVE-25579</a> : LOAD overwrite fügt statt zu überschreiben hinzu
Backport	<a href="#">HIVE-25659</a> : Metastore-Direkt-SQL-Abfragen mit IN/ (NOT IN) sollten auf der Grundlage der maximalen Anzahl von Parametern aufgeteilt werden, die von SQL DB zulässig sind
Backport	<a href="#">HIVE-20502</a> : Behebung von NPE beim Ausführen von skewjoin_mapjoin10.q, wenn Spaltenstatistiken verwendet wurden.

Typ	Beschreibung
Backport	<a href="#">HIVE-25765</a> : Die Eigenschaft <code>skip.header.line.count</code> überspringt Zeilen jedes Blocks, wenn die Dateigröße größer ist FetchOperator
Fehler	Behebung von NPE beim Einfügen in bestimmten Szenarien, wenn <code>hive.stats.column.autogather</code> und <code>hive.groupby.skewindata</code> beide aktiviert sind.
Fehler	Korrigieren Sie NPE, wenn der <code>mapred.tasktracker.expiry.interval</code> -Wert nicht gesetzt ist

## Amazon EMR 6.9.0 – Versionshinweise zu Hive

### Amazon EMR 6.9.0 – Hive-Änderungen

Typ	Beschreibung
Upgrade	Aktualisieren von Jetty auf <a href="#">9.4.48.v20220622</a>
Upgrade	Unterstützung von Hadoop 3.3.3
Funktion	Amazon-EMR-Hive-Integration mit Lake Formation für interaktive Workloads mithilfe der GCSC-API.
Funktion	Amazon-EMR-Hive-Integration mit Iceberg.
Verbesserung	Aktivieren Sie SSL in HiveServer 2, wenn die <a href="#">Verschlüsselung während der Übertragung</a> mithilfe von Amazon EMR-Sicherheitskonfigurationen aktiviert ist.
Verbesserung	Aktivieren Sie standardmäßig den optimierten Hive-EMRFS-Amazon-S3-Committer. Weitere

Typ	Beschreibung
	Informationen finden Sie unter <a href="#">Aktivierung des optimierten Hive EMRFS S3 Committers</a> .
Verbesserung	<a href="#">Fügen Sie hinzuHiveHBaseTableInputFormatV2</a> , dass nur die zugeordnete Version von erbt wird, um <a href="#">SPARK-34210 zu InputFormat reparieren</a> . Stellen Sie <code>hive.hbase.inputformat.v2</code> auf <code>true</code> ein, um es zu verwenden.
Verbesserung	Warten Sie, bis TezAM im Hintergrund mit <a href="#">hive.cli.tez.session.async</a> gestartet wird, anstatt es zu beenden und sofort neu zu starten. Verwenden Sie diese <code>hive.emr.cli.tez.session.open.timeout</code> -Option, um dieses Timeout in Sekunden festzulegen.
Verbesserung	Fügen Sie die Option <a href="#">hive.conf.restricted.list.append</a> hinzu, um kommagetrennte Konfigurationen an die bestehende eingeschränkte Konfigurationsliste <code>hive.conf.restricted.list</code> anzuhängen.
Verbesserung	Klarere Fehlermeldung, wenn die Hive-Abfrage fehlschlägt, weil der Speicherort für die Datenbank nicht definiert ist.
Backport	<a href="#">HIVE-24484</a> : Aktualisieren Sie Hadoop auf 3.3.1 und Tez auf 0.10.2
Backport	<a href="#">HIVE-22398</a> : Entfernen Sie die YARN-Warteschlangenverwaltung über ShimLoader
Backport	<a href="#">HIVE-23190: LLAP: modifizieren</a> , um das Dateisystemobjekt zu übergeben. <code>IndexCache</code> <code>TezSpillRecord</code>

Typ	Beschreibung
Backport	<a href="#">HIVE-22185</a> : HADOOP-15832 wird Probleme bei Tests verursachen, die Cluster verwenden. MiniYarn
Backport	<a href="#">HIVE-21670</a> : Ersetzen Sie mockito-all durch mockito-core -Abhängigkeit.
Backport	<a href="#">HIVE-24542</a> : Bereiten Sie Guava für Upgrades vor.
Backport	<a href="#">HIVE-23751</a> : QTest: #mkdirs() -Methode in ProxyFileSystem überschreiben, um sie nach HADOOP-16582 auszurichten.
Backport	<a href="#">HIVE-21603</a> : Java-11-Vorbereitung: Powermock-Version aktualisieren.
Backport	<a href="#">HIVE-24083</a> : hcatalog-Fehler in Hadoop 3.3.0: Authentifizierungstyp erforderlich.
Backport	<a href="#">HIVE-24282</a> : Spalten anzeigen sollte Ausgabespalten nicht sortieren, sofern nicht ausdrücklich anders angegeben.
Backport	<a href="#">HIVE-20656</a> : Sinnvolle Standardeinstellungen: Die Speicherkonfigurationen für die Kartenaggregation sind zu aggressiv.
Backport	<a href="#">HIVE-25443</a> : Arrow SerDe kann komplexe Datentypen nicht serialisieren/deserialisieren, wenn es mehr als 1024 Werte gibt
Backport	<a href="#">HIVE-19792</a> : Aktualisieren Sie Orc auf 1.5.2 und aktivieren Sie die Decimal_64-Schema-Evolutionstests.

Typ	Beschreibung
Backport	<a href="#">HIVE-20437</a> : Behandelt die Schemaentwicklung von Float, Double und Decimal.
Backport	<a href="#">HIVE-21987</a> : Hive kann Parquet int32, das mit Dezimalzahlen annotiert ist, nicht lesen.
Backport	<a href="#">HIVE-20038</a> : Bei Aktualisierungsabfragen für Tabellen ohne Buckets und partitionierte Tabellen wird NPE ausgelöst.

## Amazon EMR 6.9.0 – Bekannte Probleme mit Hive

- Bei Amazon EMR 6.6.0 bis 6.9.x haben INSERT-Abfragen mit dynamischer Partition und einer ORDER-BY- oder SORT-BY-Klausel immer zwei Reduzierungen. Dieses Problem wird durch die OSS-Änderung [HIVE-20703](#) verursacht, die die Optimierung dynamischer Sortierpartitionen einer kostenbasierten Entscheidung unterstellt. Wenn Ihr Workload keine Sortierung dynamischer Partitionen erfordert, empfehlen wir Ihnen, die `hive.optimize.sort.dynamic.partition.threshold`-Eigenschaft auf -1 zu setzen, um das neue Feature zu deaktivieren und die korrekt berechnete Anzahl von Reduzierern zu erhalten. Dieses Problem wurde behoben in OSS Hive als Teil von [HIVE-22269](#) und in Amazon EMR 6.10.0 behoben.

## Amazon EMR 6.8.0 – Versionshinweise zu Hive

### Amazon EMR 6.8.0 – Hive-Änderungen

Typ	Beschreibung
Verbesserung	Reduzieren Sie die Dateisystemaufrufe im Befehl <code>msck</code> . Leistungsverbesserungen (~15-20x auf mehr als 10 000 Partitionen)
Backport	<a href="#">HIVE-20678: HiveH sollte implementiert werden, um die Kompatibilität sicherzustellen</a> BaseTableOutputFormat HiveOutputFormat



Typ	Beschreibung
Backport	<a href="#">HIVE-21040</a> : msck listet unnötige Dateien auf der letzten Ebene des Verzeichnisbaums auf
Backport	<a href="#">HIVE-21460</a> : Das Laden von Daten, gefolgt von einer Auswahl-*-Abfrage, führt zu falschen Ergebnissen
Backport	<a href="#">HIVE-21660</a> : Falsches Ergebnis, wenn Union All und spätere Ansicht mit Explode verwendet werden
Backport	<a href="#">HIVE-22505</a> : verursacht durch eine falsche Auswahl des vektorisierten Operators ClassCastException
Backport	<a href="#">HIVE-22513</a> : Die ständige Ausbreitung von gegossenen Spalten in Filteroperationen kann zu falschen Ergebnissen führen
Backport	<a href="#">HIVE-23435</a> : Beim vollständigen Outer-Join-Ergebnis fehlen Zeilen
Backport	<a href="#">HIVE-24209</a> : Falsche Konvertierung der Suchargumente für die Operation NOT BETWEEN, wenn die Vektorisierung aktiviert ist
Backport	<a href="#">HIVE-24934</a> : In GenericUdfSQ ist keine Anmerkung erforderlich VectorizedExpressions CountCheck
Backport	<a href="#">HIVE-25278</a> : kann ungültige Transformationen mit Fensterausdrücken durchführen HiveProjectJoinTransposeRule
Backport	<a href="#">HIVE-25505</a> : Falsche Ergebnisse mit Header. skip.header.line.count, wenn die erste Zeile leer ist

Typ	Beschreibung
Backport	<a href="#">HIVE-26080</a> : Accumulo-Core auf 1.10.1 aktualisieren
Backport	<a href="#">HIVE-26235</a> : OR-Bedingung für die Binärspalte gibt ein leeres Ergebnis zurück
Fehler	Behebt mehrere Warnprotokolle für SLF4J-Bindungen in stderr während des Starts
Fehler	Behebt die Abfrage SHOW TABLE EXTENDED, die mit dem Fehler Falscher FS fehlschlägt, wenn sich Partition und Tabelle auf unterschiedlichen Dateisystemen befinden.

## Amazon EMR 6.8.0 – Bekannte Probleme mit Hive

- Bei Amazon EMR 6.6.0 bis 6.9.x haben INSERT-Abfragen mit dynamischer Partition und einer ORDER-BY- oder SORT-BY-Klausel immer zwei Reduzierungen. Dieses Problem wird durch die OSS-Änderung [HIVE-20703](#) verursacht, die die Optimierung dynamischer Sortierpartitionen einer kostenbasierten Entscheidung unterstellt. Wenn Ihr Workload keine Sortierung dynamischer Partitionen erfordert, empfehlen wir Ihnen, die `hive.optimize.sort.dynamic.partition.threshold`-Eigenschaft auf -1 zu setzen, um das neue Feature zu deaktivieren und die korrekt berechnete Anzahl von Reduzierern zu erhalten. Dieses Problem wurde behoben in OSS Hive als Teil von [HIVE-22269](#) und in Amazon EMR 6.10.0 behoben.

## Amazon EMR 6.7.0 – Versionshinweise zu Hive

### Amazon EMR 6.7.0 – Hive-Änderungen

Typ	Beschreibung
Funktion	<a href="#">Amazon EMR Hive-Integration mit LakeFormation</a>

Typ	Beschreibung
Funktion	Zusätzliche Auditprotokollierung für den für Hive EMRFS Amazon S3 optimierten Committer. Hive-Konfiguration: <code>hive.blobstore.output-committer.logging</code> , Standard: <code>false</code>
Funktion	Das Zielverzeichnis wurde beim Einfügen gelöscht und mit leerem Auswahlergebnis in eine unpartitionierte Tabelle/statische Partition überschrieben, sodass es sich ähnlich wie Hive 2.x verhält. Hive-Konfiguration: <code>hive.emr.iow.clean.target.dir</code> , Standard: <code>false</code>
Fehler	Es wurde ein zeitweiliger Abfragefehler bei der Verwendung des für Amazon S3 optimierten Hive EMRFS Committers mit Partitions-Bucket-Sortierung behoben.
Upgrade	Upgrade von Hive auf Version 3.1.3. Weitere Informationen finden Sie in den <a href="#">Versionshinweisen zu Apache Hive 3.1.3</a> .
Upgrade	<a href="#">Parquet wurde auf 1.12.2 aktualisiert.</a>
Backport	<a href="#">HIVE-20065</a> : Metastore sollte sich nicht auf Jackson 1.x verlassen
Backport	<a href="#">HIVE-20071</a> : Migrieren Sie zu Jackson 2.x und verhindern Sie die Nutzung
Backport	<a href="#">HIVE-20607</a> : TxnHandler sollte verwendet werden, um direkte SQL-Abfragen auszuführen PreparedStatement

Typ	Beschreibung
Backport	<a href="#">HIVE-20740</a> : Entfernen Sie die globale Sperre in der Methode.setConf ObjectStore
Backport	<a href="#">HIVE-20961</a> : NVL-Implementierung einstellen
Backport	<a href="#">HIVE-22059</a> : Hive-Exec-Jar enthält keine Jackson-Bibliothek (fasterxml)
Backport	<a href="#">HIVE-22351</a> : Korrigiert die falsche Verwendung von Threads in ObjectStore TestObjectStore
Backport	<a href="#">HIVE-23534</a> : NPE in #invoke beim Abfangen ohne Nachricht RetryingMetaStoreClient MetaException
Backport	<a href="#">HIVE-24048</a> : Harmonisieren Sie die Jackson-Komponenten auf Version 2.10.latest – Hive
Backport	<a href="#">HIVE-24768</a> : Verwenden Sie jackson-bom überall für den Versionsersatz
Backport	<a href="#">HIVE-24816</a> : Aktualisieren Sie Jackson aufgrund von CVE-2020-25649 auf 2.10.5.1 oder 2.11.0+
Backport	<a href="#">HIVE-25971</a> : Das Herunterfahren der Tez-Aufgabe verzögert sich, weil der zwischeng espeicherte Thread-Pool nicht geschlossen wurde
Backport	<a href="#">HIVE-26036</a> : NPE verursacht durch getMTable () in ObjectStore

## Amazon EMR 6.7.0 – Bekannte Probleme mit Hive

- Abfragen mit Fensterfunktionen für dieselbe Spalte wie Join können zu ungültigen Transformationen führen, wie in [HIVE-25278](#) gemeldet, und zu falschen Ergebnissen oder

Abfragefehlern führen. Eine Problemumgehung bestünde darin, CBO für solche Abfragen auf Abfrageebene zu deaktivieren. Das Update wird in einer Amazon-EMR-Version nach 6.7.0 verfügbar sein. Für weitere Informationen wenden Sie sich an den Support. AWS

- Bei Amazon EMR 6.6.0 bis 6.9.x haben INSERT-Abfragen mit dynamischer Partition und einer ORDER-BY- oder SORT-BY-Klausel immer zwei Reduzierungen. Dieses Problem wird durch die OSS-Änderung [HIVE-20703](#) verursacht, die die Optimierung dynamischer Sortierpartitionen einer kostenbasierten Entscheidung unterstellt. Wenn Ihr Workload keine Sortierung dynamischer Partitionen erfordert, empfehlen wir Ihnen, die `hive.optimize.sort.dynamic.partition.threshold`-Eigenschaft auf -1 zu setzen, um das neue Feature zu deaktivieren und die korrekt berechnete Anzahl von Reduzierern zu erhalten. Dieses Problem wurde behoben in OSS Hive als Teil von [HIVE-22269](#) und in Amazon EMR 6.10.0 behoben.

## Amazon EMR 6.6.0 – Versionshinweise zu Hive

### Amazon EMR 6.6.0 – Hive-Änderungen

Typ	Beschreibung
Upgrade	Aktualisieren Sie Parquet auf <a href="#">1.12.1</a>
Upgrade	Aktualisieren der Jetty-Jar-Version auf 9.4.43.v20210629
Fehler	Es wurde ein Problem behoben, das dazu führte, dass Hive auf allen Aufgaben-/Core-Knoten installiert wurde, wenn LLAP auf einem Hive-Cluster aktiviert war.
Backport	<a href="#">HIVE-25942</a> : Aktualisieren Sie Commons-io aufgrund von CVE-2021-29425 auf 2.8.0
Backport	<a href="#">HIVE-25726</a> : Erhöhen Sie die Geschwindigkeit aufgrund von CVE-2020-13936 auf 2,3
Backport	<a href="#">HIVE-25680</a> : Autorisieren Sie die <code>#get_table_meta</code> HiveMetastore Server-API, um eines

Typ	Beschreibung
	der Autorisierungsmodelle zu verwenden. HiveMetastore
Backport	<a href="#">HIVE-25554</a> : Aktualisieren Sie die Pfeilverion auf 0.15
Backport	<a href="#">HIVE-25242</a> : Die Abfrage wird mit vectorize d.adaptor = chosen extrem langsam ausgeführt
Backport	<a href="#">HIVE-25085</a> : MetaStore Clients werden nicht mehr sitzungsübergreifend gemeinsam genutzt.
Backport	<a href="#">HIVE-24827</a> : Die Hive-Aggregationsabfrage gibt falsche Ergebnisse für Nicht-Textdateien zurück.
Backport	<a href="#">HIVE-24683</a> : Hadoop23Shims sind aufgrund nicht vorhandener Pfade anfällig für NPE getFiled
Backport	<a href="#">HIVE-24656</a> : CBO schlägt bei Abfragen mit dem Wert Null bei Map- und Array-Typen fehl
Backport	<a href="#">HIVE-24556</a> : Optimiert DefaultGraphWalker für den Fall ohne Enkelkind
Backport	<a href="#">HIVE-24408</a> : Parquet auf 1.11.1 aktualisieren
Backport	<a href="#">HIVE-24391</a> : FIX-Fehler in Branch-3.1 behoben TestOrcFile
Backport	<a href="#">HIVE-24362</a> : Die AST-Baumverarbeitung ist für Bäume mit einer großen Anzahl von Knoten suboptimal
Backport	<a href="#">HIVE-24316</a> : Aktualisieren Sie ORC von 1.5.6 auf 1.5.8 in Branch-3.1

Typ	Beschreibung
Backport	<a href="#">HIVE-24307</a> : Beeline mit Eigenschaftsdatei und -e-Parameter schlägt fehl
Backport	<a href="#">HIVE-24245</a> : Vektorisiertes PTF mit Count und Distinct über der Partition, was zu falschen Ergebnissen führt.
Backport	<a href="#">HIVE-24224</a> : Das Überspringen der Kopf-/Fußzeile für Hive on Tez in komprimierter Datei wurde behoben
Backport	<a href="#">HIVE-24157</a> : Strikter Modus schlägt bei CAST-Zeitstempel fehl ↔ numerisch
Backport	<a href="#">HIVE-24113</a> : <a href="#">NPE in GenericUDF ToUnixTimeStamp</a>
Backport	<a href="#">HIVE-23987</a> : Aktualisieren von Arrow auf Version 0.11.0
Backport	<a href="#">HIVE-23972</a> : Fügen Sie dem externen LLAP-Client eine externe Client-ID hinzu
Backport	<a href="#">HIVE-23806</a> : Vermeiden Sie es, den Status der Spaltenstatistiken auf allen Partitionen zu löschen, falls das Schema erweitert wird. Dies verbessert die Laufzeit der Anweisung alter table add columns.
Backport	<a href="#">HIVE-23779</a> : Die Informationen werden in der Beeline-Konsole nicht gedruckt BasicStatsTask
Backport	<a href="#">HIVE-23306</a> : Der RESET-Befehl funktioniert nicht, wenn eine von System.GetProperty festgelegte Konfiguration vorhanden ist

Typ	Beschreibung
Backport	<a href="#">HIVE-23164</a> : Der Server wurde aufgrund von Threads, die kein Daemon sind, nicht ordnungsgemäß beendet
Backport	<a href="#">HIVE-22967</a> : Support hive.reloadable.aux.jars.path für Hive on Tez
Backport	<a href="#">HIVE-22934</a> : Interaktive Protokollzähler des Hive-Servers zum Fehlerstream
Backport	<a href="#">HIVE-22901</a> : Die Substitution von Variablen kann bei Zirkelverweisen zu OOM führen
Backport	<a href="#">HIVE-22769</a> : Falsche Abfrageergebnisse und Abfragefehler bei der Split-Generierung für komprimierte Textdateien
Backport	<a href="#">HIVE-22716</a> : Das Lesen in ist unterbrochen ByteBuffer ParquetFooterInputFromCache
Backport	<a href="#">HIVE-22648</a> : Parquet auf 1.11.0 aktualisieren
Backport	<a href="#">HIVE-22640</a> : <a href="#">Decimal64: wenn</a> der Partition spaltentyp Dezimal ist ColumnVector ClassCastException
Backport	<a href="#">HIVE-22621</a> : <a href="#">instabiler Testfall: .testSigning</a> TestLlapSignerImpl
Backport	<a href="#">HIVE-22533</a> : Behebt mögliche Sicherheitslücken in der Web-Benutzeroberfläche des LLAP-Daemons
Backport	<a href="#">HIVE-22532</a> : PTFPPD kann das Limit durch die Funktion Rank/ falsch verschieben DenseRank



Typ	Beschreibung
Backport	<a href="#">HIVE-22514</a> : könnte viel Speicher verbrauchen HiveProtoLoggingHook
Backport	<a href="#">HIVE-22476</a> : Die Hive-Datediff-Funktion lieferte inkonsistente Ergebnisse, wenn hive.fetch.task.conversion auf none gesetzt war
Backport	<a href="#">HIVE-22429</a> : Migrierte Clustertabellen mit bucketing_version 1 auf Hive 3 verwenden bucketing_version 2 für Einfügungen
Backport	<a href="#">HIVE-22412</a> : Beim Erklären wird NPE ausgelöst StatsUtils
Backport	<a href="#">HIVE-22360</a> : MultiDelimiterSerDe gibt falsche Ergebnisse in der letzten Spalte zurück, wenn die geladene Datei mehr Spalten hat als die im Tabellenschema
Backport	<a href="#">HIVE-22332</a> : Hive sollte sicherstellen, dass die Einstellungen für die Schemaentwicklung seit ORC-540 gültig sind
Backport	<a href="#">HIVE-22331</a> : unix_timestamp ohne Argument gibt den Zeitstempel in Millisekunden statt in Sekunden zurück
Backport	<a href="#">HIVE-22275</a> : OperationManager queryIdOperationbereinigt mehrere QueryIDs nicht richtig
Backport	<a href="#">HIVE-22273</a> : Die Zugriffsprüfung schlägt fehl, wenn ein temporäres Verzeichnis entfernt wird
Backport	<a href="#">HIVE-22270</a> : Aktualisieren Sie Commons-io auf 2.6

Typ	Beschreibung
Backport	<a href="#">HIVE-22241</a> : Implementieren Sie UDF zur Interpretation von Datum/Zeitstempel mithilfe seiner internen Darstellung und des gregorianisch-julianischen Hybridkalenders
Backport	<a href="#">HIVE-22241</a> : Implementieren Sie UDF zur Interpretation von Datum/Zeitstempel mithilfe seiner internen Darstellung und des gregorianisch-julianischen Hybrids
Backport	<a href="#">HIVE-22232</a> : NPE, wenn hive.order.columnalignment auf false gesetzt ist
Backport	<a href="#">HIVE-22231</a> : Hive-Abfrage mit großer Größe über Knox schlägt fehl, Broken Pipe Write ist fehlgeschlagen
Backport	<a href="#">HIVE-22221</a> : Externer Llap-Client — #getSplits muss reduziert werden LlapBaselInputFormat
Backport	<a href="#">HIVE-22208</a> : Der Spaltenname mit reserviertem Schlüsselwort wird nicht maskiert, wenn die Abfrage, die eine Verknüpfung für eine Tabelle mit Maskenspalte enthält, neu geschrieben wird
Backport	<a href="#">HIVE-22197</a> : Häufiger Merge-Join, der eine Klassenumwandlungs-Ausnahme auslöst.
Backport	<a href="#">HIVE-22170</a> : from_unixtime und unix_time stamp sollten die Zeitzone der Benutzersitzung verwenden
Backport	<a href="#">HIVE-22169</a> : Tez: SplitGenerator versucht nach Plandateien zu suchen, die für Tez nicht existieren

Typ	Beschreibung
Backport	<a href="#">HIVE-22168</a> : Entfernen Sie sehr teure Protokollierung aus dem LLap-Cache-Hotpath
Backport	<a href="#">HIVE-22161</a> : UDF: synchronisiert mit der Klasse <code>org.apache.hadoop.hive ql.udf.UDFType</code> <code>FunctionRegistry</code>
Backport	<a href="#">HIVE-22120</a> : Korrigiert falsche <code>ArrayOutOfBounds</code> Ergebnisse/Ausnahmen bei Verbindungen auf der linken äußeren Karte unter bestimmten Randbedingungen
Backport	<a href="#">HIVE-22115</a> : Verhindert die Erstellung eines Query Routing Appenders, wenn die Eigenschaft auf <code>False</code> gesetzt ist
Backport	<a href="#">HIVE-22113</a> : Verhindert das Herunterfahren von LLAP im Zusammenhang mit <code>AmReporter RuntimeException</code>
Backport	<a href="#">HIVE-22106</a> : Entfernen Sie die abfrageübergreifende Synchronisation für die Partition-Eval
Backport	<a href="#">HIVE-22099</a> : Seit HIVE-20007 können mehrere datumsbezogene UDFs julianische Datumsangaben nicht richtig verarbeiten
Backport	<a href="#">HIVE-22037</a> : HS2 sollte sich beim Herunterfahren aufgrund von OOM protokollieren
Backport	<a href="#">HIVE-21976</a> : Der Offset sollte in Calcit Null statt Null sein <code>HiveSortLimit</code>
Backport	<a href="#">HIVE-21924</a> : Textdateien teilen, auch wenn Kopf-/Fußzeile vorhanden ist

Typ	Beschreibung
Backport	<a href="#">HIVE-21913</a> : GetSplits GenericUDTF sollte Benutzernamen genauso behandeln wie LLAP
Backport	<a href="#">HIVE-21905</a> : Verbesserung der Generika FetchOperator in der gesamten Klasse
Backport	<a href="#">HIVE-21902</a> : HiveServer 2 UI: Der Jetty-Antwort-Header benötigt X-Frame-Options
Backport	<a href="#">HIVE-21888</a> : Setzen Sie hive.parquet.timesamp.skip.conversion standardmäßig auf wahr
Backport	<a href="#">HIVE-21868</a> : Vektorisieren Sie das CAST ...-FORMAT
Backport	<a href="#">LlapBaseInputFormatHIVE-21864</a> : #closeAll
Backport	<a href="#">HIVE-21863</a> : Verbessert das Vectorizer-Typ-Casting für den WHEN-Ausdruck
Backport	<a href="#">HIVE-21862</a> : ORC ppd erzeugt ein falsches Ergebnis mit Zeitstempel
Backport	<a href="#">HIVE-21846</a> : Erstelle einen Thread in TeamAm, der regelmäßig Metriken abrufft LlapDaemon
Backport	<a href="#">HIVE-21837</a> : löst eine Ausnahme aus, wenn die ausgewählte Spalte MapJoin vollständig Nullwerte enthält
Backport	<a href="#">HIVE-21834</a> : Vermeiden Sie unnötige Aufrufe, um die Filterbedingungen zu vereinfachen
Backport	<a href="#">HIVE-21832</a> : Neue Metriken zur Ermittlung der durchschnittlichen Warteschlangen-/Severierungs- und Antwortzeit

Typ	Beschreibung
Backport	<a href="#">HIVE-21827</a> : Mehrere Aufrufe durchlaufen die Methode nicht SemanticAnalyzer getObjectByName
Backport	<a href="#">HIVE-21822</a> : Machen Sie Metriken über eine neue API-Methode verfügbar LlapDaemon
Backport	<a href="#">HIVE-21818</a> : CBO: Beim Kopieren kommt es zu Metastore-Verkehr TableRelOptHiveTable
Backport	<a href="#">HIVE-21815</a> : Statistiken in der ORC-Datei werden zweimal analysiert
Backport	<a href="#">HIVE-21805</a> : 2: Verwenden Sie die schnellen APIs HiveServer ShutdownHookManager
Backport	<a href="#">HIVE-21799</a> : NullPointerException in, wenn sich der Join-Schlüssel in der DynamicPartitionPruningOptimization Aggregationsspalte befindet
Backport	<a href="#">HIVE-21794</a> : Fügen Sie materialisierte Ansichtparameter hinzu zu sqlStdAuth SafeVarNameRegexes
Backport	<a href="#">HIVE-21768</a> : JDBC: Entfernt das Standard-Union-Präfix für nicht eingeschlossene UNION-Abfragen
Backport	<a href="#">HIVE-21746</a> : ArrayIndexOutOfBoundsException während eines dynamisch partitionierten Hash-Joins mit deaktiviertem CBO
Backport	<a href="#">HIVE-21717</a> : Das Umbenennen des Verzeichnisses in der Verschiebungsaufgabe schlägt fehl.

Typ	Beschreibung
Backport	<a href="#">HIVE-21685</a> : Falsche Vereinfachung in einer Abfrage mit mehreren IN-Klauseln
Backport	<a href="#">HIVE-21681</a> : Formatiertes beschreiben zeigt falsche Informationen für mehrere Primärschlüssel
Backport	<a href="#">HIVE-21651</a> : Verschieben Sie Protobug Serde nach Hive-Exec.
Backport	<a href="#">HIVE-21619</a> : Drückt den Zeitstempeltyp ohne Genauigkeit in SQL Explain Extended
Backport	<a href="#">HIVE-21592</a> : OptimizedSql wird nicht angezeigt, wenn der Ausdruck CONCAT enthält
Backport	<a href="#">HIVE-21576</a> : Einführung von CAST ... FORMAT und einer begrenzten Liste von SQL:2016-Datetime-Formaten
Backport	<a href="#">HIVE-21573</a> : Der binäre Transport ignoriert den Prinzipal, wenn Auth auf DelegationToken gesetzt ist
Backport	<a href="#">HIVE-21550</a> : Die TestObjectStore Tests sind fehlerhaft — Eine Sperre konnte nicht innerhalb der angeforderten Zeit abgerufen werden
Backport	<a href="#">HIVE-21544</a> : Ständige Ausbreitung verfälscht Koalesce/Case/When Ausdrücke beim Falten
Backport	<a href="#">HIVE-21539</a> : GroupBy + WHERE-Klausel in derselben Spalte führt zu einem falschen Umschreiben der Abfrage

Typ	Beschreibung
Backport	<a href="#">HIVE-21538</a> : Beeline: Passwortquelle, obwohl der Konsolenleser den Verbindungsparameter nicht übergeben hat
Backport	<a href="#">HIVE-21509</a> : LLAP kann beschädigte Spaltenvektoren zwischenspeichern und ein falsches Abfrageergebnis zurückgeben
Backport	<a href="#">HIVE-21499</a> : sollte die Funktion nicht aus der Registrierung entfernen, wenn der Befehl create fehlgeschlagen ist mit AlreadyExistsException
Backport	<a href="#">HIVE-21496</a> : Bei der automatischen Dimensionierung ungeordneter Puffer kann es zu einem Überlauf kommen
Backport	<a href="#">HIVE-21468</a> : Berücksichtigung von Groß- und Kleinschreibung bei Bezeichnernamen für den JDBC-Speicherhandler
Backport	<a href="#">HIVE-21467</a> : Entfernen Sie veraltete JUnit.Framework.Assert-Importe
Backport	<a href="#">HIVE-21435</a> : LlapBaselInputFormat sollte beim Erstellen die Aufgabennummer aus der TASK_ATTEMPT_ID conf abrufen, falls vorhanden SubmitWorkRequestProto
Backport	<a href="#">HIVE-21389</a> : Bei der Hive-Distribution fehlt javax.ws.rs-api.jar nach HIVE-21247
Backport	<a href="#">HIVE-21385</a> : Erlaubt die Deaktivierung der Übertragung von Berechnungen, die nicht splittbar sind, an JDBC-Quellen

Typ	Beschreibung
Backport	<a href="#">HIVE-21383</a> : JDBC-Speicherhandler: Verwenden Sie Katalog und Schema, um Tabellen abzurufen, falls angegeben
Backport	<a href="#">HIVE-21382</a> : Optimierung der Reduzierung der Gruppierung nach Schlüsseln – Schlüssel werden in query23 nicht reduziert
Backport	<a href="#">HIVE-21362</a> : Fügt ein Eingabeformat hinzu und dient dazu, aus Protobuf-Dateien zu lesen.
Backport	<a href="#">HIVE-21340</a> : CBO: Beschneidet Spalten, die keine Schlüssel sind SemiJoin
Backport	<a href="#">HIVE-21332</a> : Löscht die nicht gesperrten Puffer statt der gesperrten
Backport	<a href="#">HIVE-21329</a> : Ungeordnete Größe des Ausgabepuffers für benutzerdefinierte Tez-Laufzeit, abhängig von der Operator-Pipeline
Backport	<a href="#">HIVE-21295</a> : <a href="#">soll das Datum unter Verwendung der Hive-Konvention</a> in eine Zeichenfolge konvertieren StorageHandler
Backport	<a href="#">HIVE-21294</a> : Vektorisierung: 1-Reducer Shuffle kann die Objekt-Hash-Funktionen überspringen
Backport	<a href="#">HIVE-21255</a> : <a href="#">Entfernen in</a> QueryConditionBuilder der JdbcStorageHandler
Backport	<a href="#">HIVE-21253</a> : Support von DB2 in JDBC StorageHandler
Backport	<a href="#">HIVE-21232</a> : LLAP: Fügt einen Anbieter für zwischengespeicherte Affinitäten hinzu



Typ	Beschreibung
Backport	<a href="#">HIVE-21214</a> : MoveTask : Verwenden Sie AttemptID anstelle der Dateigröße für die Deduplizierung von Dateien compareTempOr DuplicateFiles
Backport	<a href="#">HIVE-21184</a> : Fügen Sie einen formatierten CBO-Plan mit Erläuterungen und Erläuterungen mit Kosteninformationen hinzu
Backport	<a href="#">HIVE-21182</a> : Überspringen Sie die Einrichtung von Hive Scratch Dir während der Planung
Backport	<a href="#">HIVE-21171</a> : Überspringen Sie das Erstellen von Scratch-Verzeichnissen für Tez, wenn RPC aktiviert ist
Backport	<a href="#">HIVE-21126</a> : Abfragen auf Sitzungsebene in #getSplit zulassen LlapBaselInputFormat
Backport	<a href="#">HIVE-21107</a> : Fehler „Feld kann nicht gefunden werden“ beim dynamisch partitionierten Hash-Join
Backport	<a href="#">HIVE-21061</a> : Die CTAS-Abfrage schlägt fehl, wenn die Quelle leer ist IllegalStateException
Backport	<a href="#">HIVE-21041</a> : NPE, beim Abrufen des Schemas aus dem logischen Plan ParseException
Backport	<a href="#">HIVE-21013</a> : Die Partitionsspalte konnte in Oracle nicht gefunden JdbcStorageHandler werden
Backport	<a href="#">HIVE-21006</a> : Erweitern, um Semijoin SharedWorkOptimizer zu entfernen, wenn sich eine Möglichkeit zur Wiederverwendung ergibt

Typ	Beschreibung
Backport	<a href="#">HIVE-20992</a> : Teilen Sie die Konfiguration <code>hive.metastore.dbaccess.ssl.properties</code> in aussagekräftigere Konfigurationen auf
Backport	<a href="#">HIVE-20989</a> : JDBC — Das +-Protokoll kann den Abfragefortschritt im Ruhezustand blockieren <code>GetOperationStatus</code>
Backport	<a href="#">HIVE-20988</a> : Falsche Ergebnisse für Gruppierungsabfragen mit Primärschlüssel in mehreren Spalten
Backport	<a href="#">HIVE-20985</a> : Wenn es sich bei den Eingaben des ausgewählten Operators um temporäre Spalten handelt, kann die Vektorisierung einige davon als Ausgabe wiederverwenden
Backport	<a href="#">HIVE-20978</a> : „ <code>hive.jdbc.*</code> “ sollte Folgendes hinzufügen <code>sqlStdAuth SafeVarNameRegexes</code>
Backport	<a href="#">HIVE-20953</a> : Entfernt eine Funktion aus der Funktionsregistrierung, wenn sie beim Erstellen nicht zum Metastore hinzugefügt werden kann.
Backport	<a href="#">HIVE-20952</a> : <code>VectorizationContext.java</code> reinigen
Backport	<a href="#">HIVE-20951</a> : LLAP: Setzen Sie <code>Xms</code> immer auf 50 %
Backport	<a href="#">HIVE-20949</a> : Verbessern Sie die Schätzung der PKFK-Kardinalität bei der Raumplanung
Backport	<a href="#">HIVE-20944</a> : Statistiken werden während der Abfragekompilierung nicht validiert

Typ	Beschreibung
Backport	<a href="#">HIVE-20940</a> : Überbrückt Fälle, in denen die Typauflösung von Calcite strenger ist als bei Hive.
Backport	<a href="#">HIVE-20937</a> : Die Postgres-JDBC-Abfrage schlägt mit der Meldung „LIMIT darf nicht negativ sein“ fehl
Backport	<a href="#">HIVE-20926</a> : Der Hinweis zur Reduzierung von Semi-Joins schlägt fehl, wenn die Bloom-Filter-Einträge hoch sind oder wenn es keine Statistiken gibt
Backport	<a href="#">HIVE-20920</a> : Verwenden Sie SQL-Einschränkungen, um den Algorithmus zur Neuordnung von Joins zu verbessern
Backport	<a href="#">HIVE-20918</a> : Markierung zum Aktivieren/Deaktivieren der Übertragung von Berechnungen von Calcite in die JDBC-Verbindung
Backport	<a href="#">HIVE-20915</a> : Stellen Sie die dynamische Optimierung der Sortierpartition für HoS und MR zur Verfügung
Backport	<a href="#">HIVE-20910</a> : Das Einfügen in eine Bucket-Tabelle schlägt aufgrund der Optimierung der dynamischen Partitionssortierung fehl
Backport	<a href="#">HIVE-20899</a> : Die Keytab-URI für den LLAP YARN-Dienst ist auf die Unterstützung von HDFS beschränkt
Backport	<a href="#">HIVE-20898</a> : Für zeitbezogene Funktionen dürfen Argumente nicht in einen Typ umgewandelt werden, der keine NULL-Werte zulässt

Typ	Beschreibung
Backport	<a href="#">HIVE-20881</a> : Konstante Ausbreitung vereinfacht Prognosen zu stark
Backport	<a href="#">HIVE-20880</a> : Den Standardwert für hive.stats.filter.in.min.ratio aktualisieren
Backport	<a href="#">HIVE-20873</a> : Verwenden Sie den Murmur-Hash, um die Hash-Kollision zu reduzieren VectorHashKeyWrapperTwoLong
Backport	<a href="#">HIVE-20868</a> : SMB Join schlägt zeitweise fehl, wenn ein untergeordnetes Op-In ist TezDummyOperator getFinalOp MapRecord Processor
Backport	<a href="#">HIVE-20853</a> : Machen Sie .registerDag in der Llap-Daemon-API verfügbar ShuffleHandler
Backport	<a href="#">HIVE-20850</a> : Übertragen Sie nach Möglichkeit die Groß- und Kleinschreibung von Projektionen in Dimensionstabellen
Backport	<a href="#">HIVE-20842</a> : Die in HIVE-20660 eingeführte Logik zur Schätzung von Statistiken für Gruppen nach wurde korrigiert
Backport	<a href="#">HIVE-20839</a> : Fehler „Feld kann nicht gefunden werden“ beim dynamisch partitionierten Hash-Join
Backport	<a href="#">HIVE-20835</a> : Die Wechselwirkung zwischen Einschränkungen und dem Umschreiben von MV kann zu einer Schleife im Calcite-Planer führen

Typ	Beschreibung
Backport	<a href="#">HIVE-20834</a> : <a href="#">Hive-Einträge</a> , die auf eine zwischengespeicherte Abfrage verweisen QueryResultCache SemanticAnalyzer
Backport	<a href="#">HIVE-20830</a> : In einigen Fällen schlägt die Bestätigung der Bereichsabfrage fehl JdbcStorageHandler
Backport	<a href="#">HIVE-20829</a> : Die Bereichsteilung löst NPE aus JdbcStorageHandler
Backport	<a href="#">HIVE-20827</a> : Inkonsistente Ergebnisse für leere Arrays
Backport	<a href="#">HIVE-20826</a> : Die HiveSemiJoin Regel wurde dahingehend erweitert, dass Join + Group by auf der linken Seite in Left Semi Join umgewandelt wird
Backport	<a href="#">HIVE-20821</a> : Schreiben Sie SUM0 in die Kombination SUM + COALESCE um
Backport	<a href="#">HIVE-20815</a> : .next soll keine Ausnahme essen JdbcRecordReader
Backport	<a href="#">HIVE-20813</a> : udf to_epoch_milli muss auch Timestamp ohne Zeitzone unterstützen.
Backport	<a href="#">HIVE-20804</a> : Weitere Verbesserungen der Gruppierungsoptimierung mit Einschränkungen
Backport	<a href="#">HIVE-20792</a> : Beim Einfügen eines Zeitstempels mit Zonen werden die Daten gekürzt
Backport	<a href="#">HIVE-20788</a> : Durch die erweiterte SJ-Reduzierung werden Spalten beim Erstellen von Filtern möglicherweise falsch zurückverfolgt

Typ	Beschreibung
Backport	<a href="#">HIVE-20778</a> : Die Neuordnung von Joins wird möglicherweise nicht ausgelöst, wenn alle Joins im Plan durch die Dekorrelationslogik erstellt wurden
Backport	<a href="#">HIVE-20772</a> : CPU-Zähler pro Aufgabe in LLAP aufzeichnen
Backport	<a href="#">HIVE-20768</a> : Tumbling Window UDF hinzufügen
Backport	<a href="#">HIVE-20767</a> : Mehrere Projekte zwischen Join-Operatoren können sich auf die Neuordnung von Verknüpfungen mithilfe von Einschränkungen auswirken
Backport	<a href="#">HIVE-20762</a> : Das Bereinigungsintervall von NOTIFICATION_LOG ist auf 60 Sekunden fest codiert und ist zu klein
Backport	<a href="#">HIVE-20761</a> : Die für die Aktualisierung ausgewählte Tabelle notification_sequence hat ein Wiederholungsintervall und die Anzahl der Wiederholungen ist zu klein
Backport	<a href="#">HIVE-20751</a> Aktualisieren von Hive auf Version 0.10.0
Backport	<a href="#">HIVE-20746</a> : Schließt die Datei am Ende des Tages HiveProtoHookLogger nicht.
Backport	<a href="#">HIVE-20744</a> : Verwenden Sie SQL-Einschränkungen, um den Algorithmus zur Neuordnung von Joins zu verbessern

Typ	Beschreibung
Backport	<a href="#">HIVE-20740</a> : Entfernen Sie die globale Sperre in der Methode.setConf. ObjectStore Dieser Cherrypick portiert HIVE-20740, das für Hive 3.2 und 4.x vorgesehen ist, auf 3.1.x zurück
Backport	<a href="#">HIVE-20734</a> : Beeline: Wenn beeline-site.xml ist und Hive CLI zu Beeline umleitet, sollte es den Systembenutzernamen/das Dummy-Passwort verwenden, anstatt nach einem zu fragen
Backport	<a href="#">HIVE-20731</a> : Die Eingabe der Keystore-Datei sollte autorisiert sein JdbcStorageHandler
Backport	<a href="#">HIVE-20720</a> : Fügt dem JDBC-Handler eine Partitionsspaltenoption hinzu
Backport	<a href="#">HIVE-20719</a> : Die SELECT-Anweisung schlägt nach dem UPDATE fehl, wenn die Optimierung und Vektorisierung von hive.optimize.sort.dynamic.partition aktiviert ist
Backport	<a href="#">HIVE-20718</a> : Perf-CLI-Treiber mit Einschränkungen hinzugefügt
Backport	<a href="#">HIVE-20716</a> : Setzen Sie den Standardwert für hive.cbo.stats.correlated.multi.key.join auf true
Backport	<a href="#">HIVE-20712: sollte tiefe Fälle extrahieren</a> HivePointLookupOptimizer
Backport	<a href="#">HIVE-20710</a> : Konstantes Falten erzeugt möglicherweise keine Nullkonstanten ohne Typen
Backport	<a href="#">HIVE-20706</a> : external_jdbc_table2.q schlägt zeitweise fehl

Typ	Beschreibung
Backport	<a href="#">HIVE-20704</a> : Erweiterung zur Unterstützung anderer Funktionen HivePreFilteringRule
Backport	<a href="#">HIVE-20703</a> : Unterstellen Sie die Optimierung dynamischer Sortierpartitionen einer kostenbasierten Entscheidung
Backport	<a href="#">HIVE-20702</a> : Berücksichtigen Sie den Mehraufwand, der durch datenstrukturabhängige Schätzungen bei der Mapjoin-Auswahl entsteht
Backport	<a href="#">HIVE-20692</a> : Aktiviert die Faltung von NOT x IS (NOT) [TRUE FALSE] -Ausdrücken
Backport	HIVE-20691: Korrigieren Sie <a href="#">org.apache.hadoop.hive.cli</a> . TestMiniLlapCliDriver. testCliDriver[cttl]
Backport	<a href="#">HIVE-20682</a> : Die asynchrone Abfrageausführung kann möglicherweise fehlschlagen, wenn Shared SessionHive vom Master-Thread geschlossen wird
Backport	<a href="#">HIVE-20676</a> : HiveServer 2: PrivilegeSynchronizer ist nicht auf den Daemon-Status gesetzt
Backport	<a href="#">HIVE-20660</a> : Die Schätzung der Gruppierung nach Statistiken könnte verbessert werden, indem die Gesamtzahl der Zeilen an die Quelltable gebunden wird
Backport	<a href="#">HIVE-20652</a> : JdbcStorageHandler Push-Join von zwei verschiedenen Datenquellen zum JDBC-Treiber



Typ	Beschreibung
Backport	<a href="#">HIVE-20651</a> : JdbcStorageHandler Das Passwort sollte verschlüsselt sein
Backport	<a href="#">HIVE-20649</a> : LLAP-fähiger Speichermanager für Orc-Writer
Backport	<a href="#">HIVE-20648</a> : LLAP: Von Operatoren gruppierte Vektoren sollten Speicher pro Executor verwenden
Backport	<a href="#">HIVE-20646</a> : Die Partitionsfilterbedingung wird nicht auf die Metastore-Abfrage übertragen, wenn sie IS NOT NULL hat
Backport	<a href="#">HIVE-20644</a> : Vermeiden Sie es, vertrauliche Informationen durch eine Hive Laufzeit-Ausnahme preiszugeben
Backport	<a href="#">HIVE-20636</a> : Verbessert die Schätzung der Anzahl der Nullwerte nach dem Outer-Join
Backport	<a href="#">HIVE-20632</a> : Die Abfrage mit der UDF get_splits schlägt fehl, wenn eine materialisierte Ansicht für die abgefragte Tabelle erstellt wird
Backport	<a href="#">HIVE-20627</a> : Gleichzeitige asynchrone Abfragen schlagen zeitweise fehl und verursachen Speicherverlust LockException
Backport	<a href="#">HIVE-20623</a> : Gemeinsames Arbeiten: Erweitert die gemeinsame Nutzung von Map-Join-Cache-Einträgen in LLAP
Backport	<a href="#">MultiDelimitSerDe HIVE-20619</a> HiveServer: Standardmäßig in 2 einschließen

Typ	Beschreibung
Backport	<a href="#">HIVE-20618</a> : Beim Zusammenführen kann die Auswahl für Tabellen ohne Buckets ausgewählt werden BucketMapJoin
Backport	<a href="#">HIVE-20617</a> : Korrigiert den Typ der Konstanten in IN-Ausdrücken, sodass sie den richtigen Typ haben
Backport	<a href="#">HIVE-20612</a> : Erstellen Sie ein neues Join-Korrelationsflag für mehrere Schlüssel für CBO
Backport	<a href="#">HIVE-20603</a> : Fehler „Falsches FS“ beim Einfügen in eine Partition nach dem Ändern des Dateisystems mit Tabellenposition
Backport	<a href="#">HIVE-20601</a> : Null im ALTER_PARTITION-Ereignis in EnvironmentContext DbNotificationListener
Backport	<a href="#">HIVE-20583</a> : Verwenden Sie den kanonischen Hostnamen nur für die Kerberos-Authentifizierung in HiveConnection
Backport	<a href="#">HIVE-20582</a> : Machen Sie hflush in der Hive-Proto-Protokollierung konfigurierbar
Backport	<a href="#">HIVE-20563</a> : Vektorisierung: Der CASE WHEN-Ausdruck schlägt fehl, wenn der THEN/ELSE-Typ und der Ergebnistyp unterschiedlich sind
Backport	<a href="#">HIVE-20558</a> : Ändern Sie die Standardeinstellung von hive.hashtable.key.count.adjustment auf 0,99
Backport	<a href="#">HIVE-20552</a> : LogicalPlan Schema schneller abrufen

Typ	Beschreibung
Backport	<a href="#">HIVE-20550</a> : Wechseln Sie zu WebHCat, um Beeline zum Senden von Hive-Abfragen zu verwenden
Backport	<a href="#">HIVE-20537</a> : In mehreren Spalten werden Schätzungen mit unkorrelierten Spalten verknüpft, die sich in CBO und Hive unterscheiden
Backport	<a href="#">HIVE-20524</a> : Die Überprüfung der Schemaentwicklung ist beim Übergang von Hive Version 2 zu Version 3 für ALTER TABLE VARCHAR zu DECIMAL unterbrochen
Backport	<a href="#">HIVE-20522</a> : HiveFilterSetOpTransposeRule Kann aufgrund der NULL-Zulässigkeit von Feldern einen Assertionsfehler auslösen
Backport	<a href="#">HIVE-20521</a> : HS2 doas=True hat ein Berechtigungsproblem mit hadoop.tmp.dir, mit MR- und S3A-Dateisystemen
Backport	<a href="#">HIVE-20515</a> : Leere Abfrageergebnisse bei Verwendung von Ergebnis-Cache und Abfrage-Temp-Verzeichnis, Ergebnis-Cache-Verzeichnis in verschiedenen Dateisystemen
Backport	<a href="#">HIVE-20508</a> : Hive unterstützt keine Benutzernamen vom Typ „user @realm“
Backport	<a href="#">HIVE-20507</a> : Beeline: Fügen Sie einen Utility-Befehl hinzu, um alle Uris aus beeline-site.xml abzurufen
Backport	<a href="#">HIVE-20505</a> : Aktualisieren Sie org.openjdk.jmh:jmh-core auf 1.21

Typ	Beschreibung
Backport	<a href="#">HIVE-20503</a> : Verwenden Sie bei der Mapjoin-Auswahl Schätzungen, die die Datenstruktur berücksichtigen
Backport	<a href="#">HIVE-20498</a> : Support den Datumstyp für die automatische Erfassung von Spaltenstatistiken
Backport	<a href="#">HIVE-20496: Vektorisierung: Vektorisiertes PTF</a> IllegalStateException
Backport	<a href="#">HIVE-20494</a> : GenericUDF ist RestrictInformationSchema nach HIVE-19440 kaputt
Backport	<a href="#">HIVE-20477: wird nicht angezeigt, wenn der Ausdruck INs enthält</a> OptimizedSql
Backport	<a href="#">HIVE-20467</a> : Wenn nicht vorhanden/wenn vorhanden, beim Erstellen/Löschen eines Ressourcenplans zulassen
Backport	<a href="#">HIVE-20462</a> : „CREATE VIEW IF NOT EXISTS“ schlägt fehl, wenn die Ansicht bereits existiert
Backport	<a href="#">HIVE-20455</a> : Protokollausgabe von security.authorization.PrivilegeSynchronizer. ausführen
Backport	<a href="#">HIVE-20439</a> : Verwenden Sie das überhöhte Speicherlimit bei der Join-Auswahl für Llap
Backport	<a href="#">HIVE-20433</a> : Die implizite Konvertierung von Zeichenketten in Zeitstempel ist langsam
Backport	<a href="#">HIVE-20432</a> : Schreiben Sie BETWEEN in IN für Integer-Typen zur Schätzung von Statistiken um

Typ	Beschreibung
Backport	<a href="#">HIVE-20423</a> : Stellen Sie NULLS LAST als Standard-Nullreihenfolge ein
Backport	<a href="#">HIVE-20418</a> : LLAP IO verarbeitet ORC-Dateien, bei denen der Zeilenindex deaktiviert ist, möglicherweise nicht korrekt für Abfragen, bei denen keine Spalten ausgewählt sind
Backport	<a href="#">HIVE-20412</a> : NPE ist da HiveMetaHook
Backport	<a href="#">HIVE-20406</a> : Nested Coalesce liefert falsche Ergebnisse
Backport	<a href="#">HIVE-20399</a> : CTAS mit einem benutzerdefinierten Tabellenspeicherort, der nicht vollständig qualifiziert ist, schlägt für MM-Tabellen fehl
Backport	<a href="#">HIVE-20393</a> : Semijoin-Reduzierung: ForDPP verhält sich inkonsistent markSemiJoin
Backport	<a href="#">HIVE-20391</a> : kann bei der Zerlegung der Aggregatfunktion auf einen falschen Rückgabebetyp schließen HiveAggregateReduceFunctionRule
Backport	<a href="#">HIVE-20383</a> : Ungültiger Warteschlangenname und Synchronisationsprobleme im Hive Proto Events Hook.
Backport	<a href="#">HIVE-20367</a> : Vektorisierung: Support Streaming für PTF AVG, MAX, MIN, SUM
Backport	<a href="#">HIVE-20366</a> : Die Schätzungen der Statistiken von TPC-DS query78 sind für seinen Nullfilter deaktiviert

Typ	Beschreibung
Backport	<a href="#">HIVE-20364</a> : Standardeinstellung für hive.map.aggr.hash.min.reduction aktualisieren
Backport	<a href="#">HIVE-20352</a> : Vektorisierung: Gruppierungsfunktion Support
Backport	<a href="#">HIVE-20347</a> : hive.optimize.sort.dynamic.partition sollte mit partitionierten CTAS und MV funktionieren
Backport	<a href="#">HIVE-20345</a> : Die Drop-Datenbank kann hängen bleiben, wenn die Tabellen bei einem anderen Aufruf gelöscht werden
Backport	<a href="#">HIVE-20343</a> : Hive 3: CTAS respektiert transactional_properties nicht
Backport	<a href="#">HIVE-20340</a> : Druid benötigt explizite CASTs von Timestamp zu STRING, wenn die Ausgabe der Timestamp-Funktion als Zeichenfolge verwendet wird
Backport	<a href="#">HIVE-20339</a> : Vektorisierung: Hebt eine unnötige Einschränkung auf, die dazu führt, dass einige PTFs mit RANK nicht vektorisiert werden
Backport	<a href="#">HIVE-20337</a> :: CachedStore Expr füllt die Partitionsliste nicht korrekt getPartitionsBy
Backport	<a href="#">HIVE-20336</a> : Maskierungs- und Filterrichtlinien für materialisierte Ansichten
Backport	<a href="#">HIVE-20326</a> : Erstellen Sie Einschränkungen mit RELY als Standard anstelle von NO RELY

Typ	Beschreibung
Backport	<a href="#">HIVE-20321: Vektorisierung: Reduziert die Speichergröße von 1 Spalte auf &lt;1 VectorHas hKeyWrapper CacheLine</a>
Backport	<a href="#">HIVE-20320</a> : Schalten Sie das Flag <code>hive.optimize.remove.sq_count_check</code> ein
Backport	<a href="#">HIVE-20315</a> : Vektorisierung: Behebt mehr Probleme mit NULL-/falschen Ergebnissen und vermeidet unnötige Umwandlungen/Konvertierungen
Backport	<a href="#">HIVE-20314</a> : Schließt das Bereinigen von Partitionen in das Umschreiben materialisierter Ansichten ein
Backport	<a href="#">HIVE-20312</a> : Erlaube den Arrow-Clients, ihre eigenen zu verwenden mit <code>BufferAllocator LlapOutputFormatService</code>
Backport	<a href="#">HIVE-20302</a> : LLAP: Die nicht vektorisierte Ausführung in IO ignoriert virtuelle Spalten, einschließlich <code>ROW__ID</code>
Backport	<a href="#">HIVE-20300</a> : <code>VectorFileSinkArrowOperator</code>
Backport	<a href="#">HIVE-20299</a> : Mögliches Rennen im LLAP-Unterezeichner-Unit-Test
Backport	<a href="#">HIVE-20296</a> : Verbessern, um aus komplexeren Kontexten extrahieren <code>HivePointLookupOptimizerRule</code> zu können
Backport	<a href="#">HIVE-20294</a> : Vektorisierung: Probleme mit NULL/Falschen Ergebnissen in COALESCE/ELT behoben

Typ	Beschreibung
Backport	<a href="#">HIVE-20292</a> : Schlechte Join-Reihenfolge in tpcds query93 mit definierter primärer Einschränkung
Backport	<a href="#">HIVE-20290</a> : Lazy initialisieren ArrowColumnarBatchSerDe , sodass währenddessen keine Puffer zugewiesen werden GetSplits
Backport	<a href="#">HIVE-20281</a> : SharedWorkOptimizer schlägt fehl, wenn „der Inhalt des Operator-Cache und der tatsächliche Plan unterschiedlich sind“
Backport	<a href="#">HIVE-20277</a> : Vektorisierung: Groß-/Kleininschreibung, die BOOLEAN zurückgeben, werden für FILTER nicht unterstützt
Backport	<a href="#">HIVE-20267</a> : Erweiterung der WebUI um ein Formular zur dynamischen Konfiguration von Protokollebenen
Backport	<a href="#">HIVE-20263</a> : Tippfehler in der Variablen HiveReduceExpressionsWithStatsRule
Backport	<a href="#">HIVE-20260</a> : Der NDV einer Spalte sollte nicht skaliert werden, wenn die Zeilenanzahl durch einen Filter in einer anderen Spalte geändert wird
Backport	<a href="#">HIVE-20252</a> : Reduzierung von Semijoins: Zyklen aufgrund von Semi-Join-Verzweigungen können unentdeckt bleiben, wenn der Seite der kleinen Tabelle ein Map-Join vorgeschaltet ist.
Backport	<a href="#">HIVE-20245</a> : Vektorisierung: Probleme mit NULL/Falschen Ergebnissen zwischen BETWEEN/IN behoben



Typ	Beschreibung
Backport	<a href="#">HIVE-20241</a> : Support die Partitionierungsspezifikation in CTAS-Anweisungen
Backport	<a href="#">HIVE-20240</a> : Semijoin Reduction: Verwenden Sie die lokale Variable, um den Zustand der externen Tabelle zu überprüfen
Backport	<a href="#">HIVE-20226</a> : HMS löst eine Ausnahme aus, wenn die Anfrage max_Events getNextNotification die max_rows der Tabelle überschreitet
Backport	<a href="#">HIVE-20225</a> : zur Unterstützung des Teradata-Binärformats SerDe
Backport	<a href="#">HIVE-20213</a> : Aktualisieren Sie Calcite auf 1.17.0
Backport	<a href="#">HIVE-20212</a> : Hiveserver2 gibt im HTTP-Modus die Metrik Default.General.OPEN_CONNECTIONS falsch aus
Backport	<a href="#">HIVE-20210</a> : Der Simple Fetch-Optimizer sollte dazu führen, dass der Filter für Spalten, die nicht partitioniert sind, und die Konvertierung minimal ist MapReduce
Backport	<a href="#">HIVE-20209</a> : Die Metastore-Verbindung schlägt beim ersten Versuch im Repl-Dump fehl
Backport	<a href="#">HIVE-20207</a> : Vektorisierung: NULL-/ Probleme mit falschen Ergebnissen beim Filtern/Vergleichen behoben
Backport	<a href="#">HIVE-20204</a> : Typkonvertierung bei IN

Typ	Beschreibung
Backport	<a href="#">SerDe HIVE-20203</a> : Bei Arrow gibt es ein DirectByteBuffer
Backport	<a href="#">HIVE-20197</a> : Vektorisierung: DECIMAL_64-Tests hinzugefügt, Datums-/Interval-/Zeitstempel-Arithmetik hinzugefügt und weitere GROUP BY-Aggregation hinzugefügt
Backport	<a href="#">HIVE-20193</a> : CBOInfo ist im Explain Plan Json nicht vorhanden
Backport	<a href="#">HIVE-20192</a> : HS2 mit eingebettetem Metastore gibt JDO-Objekte preis PersistenceManager
Backport	<a href="#">HIVE-20183</a> : Das Einfügen aus einer Bucket-Tabelle kann zu Datenverlust führen, wenn die Quelltable einen leeren Bucket enthält
Backport	<a href="#">HIVE-20177</a> : Vektorisierung: Reduzieren Sie die Zuweisung im Streaming-Modus KeyWrapper GroupBy
Backport	<a href="#">HIVE-20174</a> : Vektorisierung: Behebung von Problemen mit NULL-/falschen Ergebnissen in GROUP BY-Aggregationsfunktionen
Backport	<a href="#">HIVE-20172</a> : Beim Versuch, eine Verbindung zum Remote-Metastore StatsUpdater herzustellen, ist eine GSS-Ausnahme aufgetreten
Backport	<a href="#">HIVE-20153</a> : Count und Sum UDF verbrauchen mehr Speicher in Hive 2+
Backport	<a href="#">HIVE-20152</a> : setzt den Datenbankstatus zurück, wenn der Repl-Dump fehlschlägt, sodass die Tabelle umbenannt werden kann

Typ	Beschreibung
Backport	<a href="#">TestHiveCli HIVE-20149</a> : Fehlgeschlagen/Zeitlimit überschritten
Backport	<a href="#">HIVE-20130</a> : Bessere Protokollierung für den Informationsschema-Synchronizer
Backport	<a href="#">HIVE-20129</a> : Kehren Sie zur positionsbasierten Schemaentwicklung für Orc-Tabellen zurück
Backport	<a href="#">HIVE-2018</a> SessionStateUserAuthenticator:..getGroupNames
Backport	<a href="#">HIVE-20116</a> : <a href="#">verwendet den übergeordneten</a> TezTask Logger
Backport	<a href="#">HIVE-20115</a> : Acid-Tabellen sollten zur Analyse keinen Fußzeilenscan verwenden
Backport	<a href="#">HIVE-20103</a> : WM: Aggregieren Sie DAG-Zähler nur, wenn mindestens einer verwendet wird
Backport	<a href="#">HIVE-20101</a> : BloomkFilter: Vermeiden Sie es, die lokalen Byte [] -Arrays vollständig zu verwenden
Backport	<a href="#">HIVE-20100</a> : OpTraits : Die Auswahl von Optraits sollte aufhören, wenn eine Nichtübereinstimmung festgestellt wird
Backport	<a href="#">HIVE-20098</a> : Statistik: NPE beim Abrufen von Partitionsstatistiken für Datumsspalten
Backport	<a href="#">HIVE-20095</a> : Das Feature zur Übertragung von Berechnungen auf externe JDBC-Tabellen wurde behoben

Typ	Beschreibung
Backport	<a href="#">HIVE-20093</a> :: Mit Netty für die Buchhaltung verwenden LlapOutputFomatService ArrowBuf
Backport	<a href="#">HIVE-20090</a> : Erweitern Sie die Entwicklung von Semi-Join-Reduktionsfiltern, um neue Möglichkeiten zu entdecken
Backport	<a href="#">HIVE-20088</a> : Der Pfad zum Speicherort der Beeline-Konfiguration wurde falsch zusammengestellt
Backport	<a href="#">HIVE-20082</a> : Bei der Konvertierung in eine Zeichenfolge wird die HiveDecimal Dezimalzahl nicht korrekt formatiert
Backport	<a href="#">HIVE-20069</a> : Korrektur der Reoptimierung bei der DPP- und Semijoin-Optimierung
Backport	<a href="#">HIVE-20051</a> : Autorisierung für temporäre Tabellen überspringen
Backport	<a href="#">HIVE-20044</a> : Arrow Serde sollte Char-Werte korrekt auffüllen und leere Zeichenketten korrekt behandeln
Backport	<a href="#">HIVE-20028</a> : Die Metastore-Client-Cache-Konfiguration wird falsch verwendet
Backport	<a href="#">HIVE-20025</a> : Bereinigung von Ereignisdateien, die erstellt wurden von HiveProtoLoggingHook
Backport	<a href="#">HIVE-20020</a> : Hive contrib jar sollte nicht in lib sein
Backport	<a href="#">HIVE-20013</a> : Fügt dem Datumstyp für die Funktion to_date eine implizite Umwandlung hinzu

Typ	Beschreibung
Backport	<a href="#">HIVE-20011</a> : Verlassen Sie den Append-Modus im Proto-Logging-Hook
Backport	<a href="#">HIVE-20005</a> : acid_table_stats, acid_no_buckets usw. – Änderung der Abfrageergebnisse in der Filiale
Backport	<a href="#">HIVE-20004</a> : Falsche Skala, die von 64 verwendet wird, führt zu falschen Ergebnissen ConvertDecimal ToDecimal
Backport	<a href="#">HIVE-19995</a> : Aggregierter Zeilenverkehr für ACID-Tabellen
Backport	<a href="#">HIVE-19993</a> : Die Verwendung eines Tabellena alias, der auch als Spaltenname erscheint, ist nicht möglich
Backport	<a href="#">HIVE-19992: Vektorisierung: Fortsetzung</a> von HIVE-19951 --> Aufruf hinzufügen zu. SchemaEvolution isOnlyImplicitKonvertierung, um kodierte LLAP-I/O für ORC nur zu deaktivieren, wenn die Datentypkonvertierung nicht implizit ist
Backport	<a href="#">HIVE-19989</a> : Metastore verwendet einen falschen Anwendungsnamen für HADOOP2-Metriken
Backport	<a href="#">HIVE-19981</a> : Verwaltung von Tabellen, die vom HiveStrictManagedMigration Hilfsprogramm in externe Tabellen konvertiert wurden, sollten so eingestellt werden, dass sie Daten löschen, wenn die Tabelle gelöscht wird
Backport	<a href="#">HIVE-19967</a> : SMB-Beitritt: Benötige Optraits für PTFOperator ala GBY Op

Typ	Beschreibung
Backport	<a href="#">HIVE-19935</a> : Die Hive WM-Sitzung wurde beendet: Die Anzahl der LLAP-Aufgaben konnte nicht aktualisiert werden
Backport	<a href="#">HIVE-19924</a> : Taggt distcp-Jobs, die von Repl Load ausgeführt werden
Backport	<a href="#">HIVE-19891</a> : Das Einfügen in externe Tabellen mit benutzerdefinierten Partitionsverzeichnissen kann zu Datenverlust führen
Backport	<a href="#">HIVE-19850</a> : Das dynamische Bereinigen von Partitionen in Tez führt zu dem Fehler „Keine Arbeit für Tablescan gefunden“
Backport	<a href="#">HIVE-19806</a> : Sortiert die Qtests-Ausgabe, um unscharfe Testergebnisse zu vermeiden
Backport	<a href="#">HIVE-19770</a> : Support für CBO für Abfragen mit mehreren gleichen Spalten in Select
Backport	<a href="#">HIVE-19769</a> : Erstellen Sie spezielle Objekte für Datenbank- und Tabellennamen
Backport	<a href="#">HIVE-19765</a> : Fügen Sie Parquet-spezifische Tests hinzu BlobstoreCliDriver
Backport	<a href="#">HIVE-19759</a> : Flockiger Test: # TestRpc testServerPort
Backport	<a href="#">HIVE-19711</a> : Faktorwechsel vom Hive-Schematool
Backport	<a href="#">HIVE-19701</a> : muss nicht synchronisiert werden getDelegationToken FromMetaStore

Typ	Beschreibung
Backport	<a href="#">HIVE-19694</a> : Die Anweisung „Materialisierte Ansicht erstellen“ sollte vor der Ausführung der SQL-Anweisung von MV nach MV-Namenskonflikten suchen.
Backport	<a href="#">HIVE-19674</a> : Gruppierung nach Dezimalkonstanten wird in die Druid-Tabelle übertragen
Backport	<a href="#">HIVE-19668</a> : Über 30% des Heaps wurden durch doppelte org.antlr.runtime verschwendet. CommonToken's und doppelte Zeichenketten
Backport	<a href="#">HIVE-19663</a> : Faktorwechsel der Generierung von LLAP IO-Berichten
Backport	<a href="#">HIVE-19661</a> : Schalten Sie Hive-UDFs so um, dass sie die Re2J-Regex-Engine verwenden
Backport	<a href="#">HIVE-19628</a> : Mögliches NPE beim LLAP-Test signieren
Backport	<a href="#">HIVE-19568</a> : Aktiv/Passives HS2 HA: Direkte Verbindung zur passiven HS2-Instanz nicht zulassen
Backport	<a href="#">HIVE-19564</a> : Vektorisierung: Behebung von Problemen mit NULL/Falschen Ergebnissen in Arithmetik
Backport	<a href="#">HIVE-19552</a> : Aktivieren Sie #druidkafkamini_ba sic .q TestMiniDruidKafkaCliDriver
Backport	<a href="#">HIVE-19432</a> : GetTablesOperation ist zu langsam, wenn der Hive zu viele Datenbanken und Tabellen hat

Typ	Beschreibung
Backport	<a href="#">HIVE-19360</a> : CBO: Fügt dem Objekt ein „OptimizedSQL“ hinzu QueryPlan
Backport	<a href="#">HIVE-19326</a> : automatisches Sammeln von Statistiken: falsche Aggregation bei UNION-Abfragen
Backport	<a href="#">TestJdbcWithHIVE-19313</a> TokenStoreNoDoAs: DB-Tests schlagen fehl
Backport	<a href="#">HIVE-19285</a> : Fügen Sie Protokolle zu den Unterklassen von hinzu MetaDataOperation
Backport	<a href="#">HIVE-19235</a> : Goldene Dateien für Minimr-Tests aktualisieren
Backport	<a href="#">HIVE-19104</a> : Wenn der Test mit Retry gestartet MetaStore wird, sollten die Instanzen unabhängig sein
Backport	<a href="#">HIVE-18986</a> : Beim Umbenennen von Tabellen wird java.lang ausgeführt. StackOverflowError in DataNucleus, wenn die Tabelle eine große Anzahl von Spalten enthält
Backport	<a href="#">HIVE-18920</a> : CBO: Initialisieren Sie die Janino-Anbieter vor der ersten Abfrage
Backport	<a href="#">HIVE-18873</a> : Das unbeaufsichtigte Überspringen des Prädikats-Pushdowns für MR kann dazu führen, dass Speicherhandler ein falsches Ergebnis ausgeben HiveInputFormat
Backport	<a href="#">HIVE-18871</a> : Fehler bei der Ausführung von Hive on Tez, weil hive.aux.jars.path auf hdfs:// gesetzt wurde



Typ	Beschreibung
Backport	<a href="#">HIVE-18725</a> : Verbessert die Fehlerbehandlung für Unterabfragen, wenn ein falscher Spaltenvektorschlüssel vorhanden ist
Backport	<a href="#">HIVE-18696</a> : Die Partitionsordner werden in der Methode <code>.add_partitions_core</code> möglicherweise nicht richtig bereinigt, wenn HiveMetaStore
Backport	<a href="#">HIVE-18453</a> : ACID: Fügen Sie die Syntax „CREATE TRANSACTIONAL TABLE“ hinzu, um die ACID ORC- und Parquet-Unterstützung zu vereinheitlichen
Backport	<a href="#">HIVE-18201</a> : Deaktiviert XPROD_EDGE für <code>sq_count_chec</code>
Backport	<a href="#">HIVE-18140</a> : Statistiken für partitionierte Tabellen können in gemischten Groß- und Kleinschreibung bei Basisstatistiken schief gehen
Backport	<a href="#">HIVE-17921</a> : Die Aggregation mit Struct in LLAP führt zu einem falschen Ergebnis
Backport	<a href="#">HIVE-17896</a> : TopNKey: Erstellen Sie einen eigenständigen vektorisierbaren TopNKey-Operator
Backport	<a href="#">HIVE-17840</a> : HiveMetaStore erzeugt eine Ausnahme, wenn <code>TransactionalListeners.NotifyEvent</code> fehlschlägt
Backport	<a href="#">HIVE-17043</a> : Entfernen Sie nicht eindeutige Spalten aus der Gruppierung nach Schlüssel, falls sie später nicht referenziert werden

Typ	Beschreibung
Backport	<a href="#">HIVE-17040</a> : Schließen Sie sich der Eliminierung bei Vorhandensein einer FK-Beziehung an
Backport	<a href="#">HIVE-16839</a> : Unsymmetrische Aufrufe von OpenTransaction/CommitTransaction bei gleichzeitigem Ändern derselben Partition
Backport	<a href="#">HIVE-16100</a> : Der Optimierer für dynamisch e sortierte Partitionen verliert gleichrangige Operatoren
Backport	<a href="#">StackOverflowError HIVE-15956</a> : wenn viele Partitionen gelöscht werden
Backport	<a href="#">HIVE-15177</a> : Die Authentifizierung mit Hive schlägt fehl, wenn der Kerberos-Authentifizierungstyp auf FromSubject gesetzt ist und der Principal _HOST enthält
Backport	<a href="#">HIVE-14898</a> : HS2 sollte den Callstack bei einem leeren Auth-Header-Fehler nicht protokollieren
Backport	<a href="#">HIVE-14493</a> : Partitionierungsunterstützung für materialisierte Ansichten
Backport	<a href="#">HIVE-14431</a> : Erkennt COALESCE als CASE
Backport	<a href="#">HIVE-13457</a> : Erstellen Sie HS2-REST-API-Endpunkte zur Überwachung von Informationen
Backport	<a href="#">HIVE-12342</a> : Setzen Sie den Standardwert von hive.optimize.index.filter auf true
Backport	<a href="#">HIVE-10296</a> : Es wurde eine Cast-Ausnahme beobachtet, wenn Hive eine Multi-Join-Abfrage im Metastore ausführt

Typ	Beschreibung
Backport	<a href="#">HIVE-6980</a> : Die Tabelle mithilfe von Direct SQL löschen

## Amazon EMR 6.6.0 – Änderungen der Hive-Konfiguration

- Im Rahmen der OSS-Änderung [HIVE-20703](#) wurde die Eigenschaft `hive.optimize.sort.dynamic.partition` zum Sortieren dynamischer Partitionen durch `hive.optimize.sort.dynamic.partition.threshold` versetzt.

Die `hive.optimize.sort.dynamic.partition.threshold`-Konfiguration hat die folgenden möglichen Werte:

Wert	Beschreibung
0 (Standard)	Macht die Optimierung zur Sortierung dynamischer Partitionen zu einer kostenbasierten Entscheidung, wenn ORC-Dateien verwendet werden. Die maximal zulässige Anzahl von Schreibzugriffen in INSERT-Abfragen wird auf der Grundlage von (Executor-/Container-Speicher) * (Prozentsatz des von Orc belegten Speichers) geteilt durch den maximalen Speicher (Stripe-Größe) berechnet, der von einem einzelnen Writer belegt wird.
-1	Deaktiviert die Optimierung zur vollständigen Sortierung dynamischer Partitionen.
1	Aktiviert die globale Sortierung dynamischer Partitionen. Dadurch bleibt für jeden Partitionswert im Reducer nur ein Record Writer geöffnet, wodurch der Speicherdruck auf den Reducern reduziert wird.
2 (oder eine größere Ganzzahl)	Weist Hive an, die angegebene Ganzzahl als Schwellenwert für die maximale Anzahl von Autoren zu verwenden.

## Amazon EMR 6.6.0 – Bekannte Probleme mit Hive

- Abfragen mit Fensterfunktionen für dieselbe Spalte wie Join können zu ungültigen Transformationen führen, wie in [HIVE-25278](#) gemeldet, und zu falschen Ergebnissen oder Abfragefehlern führen. Um dieses Problem zu umgehen, können Sie CBO für solche Abfragen auf Abfrageebene deaktivieren. Wenden Sie sich für AWS weitere Informationen an den Support.
- Amazon EMR 6.6.0 enthält die Hive-Softwareversion 3.1.2. Hive 3.1.2 führt ein Feature ein, die Textdateien aufteilt, wenn sie eine Kopf- und Fußzeile enthalten ([HIVE-21924](#)). Der Apache Tez App Master liest jede Ihrer Dateien, um Versatzpunkte im Datenbereich zu bestimmen. Diese Verhaltensweisen zusammen können sich negativ auf die Leistung auswirken, wenn Ihre Abfragen eine große Anzahl kleiner Textdateien lesen. Um dieses Problem zu umgehen, können Sie `CombineHiveInputFormat` und die maximale Teilungsgröße verwenden und anpassen, indem Sie die folgenden Eigenschaften konfigurieren:

```
SET hive.tez.input.format=org.apache.hadoop.hive ql.io.CombineHiveInputFormat;  
SET mapreduce.input.fileinputformat.split.maxsize=16777216;
```

- Bei Amazon EMR 6.6.0 bis 6.9.x haben INSERT-Abfragen mit dynamischer Partition und einer ORDER-BY- oder SORT-BY-Klausel immer zwei Reduzierungen. Dieses Problem wird durch die OSS-Änderung [HIVE-20703](#) verursacht, die die Optimierung dynamischer Sortierpartitionen einer kostenbasierten Entscheidung unterstellt. Wenn Ihr Workload keine Sortierung dynamischer Partitionen erfordert, empfehlen wir Ihnen, die `hive.optimize.sort.dynamic.partition.threshold`-Eigenschaft auf -1 zu setzen, um das neue Feature zu deaktivieren und die korrekt berechnete Anzahl von Reduzierern zu erhalten. Dieses Problem wurde behoben in OSS Hive als Teil von [HIVE-22269](#) und in Amazon EMR 6.10.0 behoben.

# Hudi

[Apache Hudi](#) ist ein Open-Source-Datenverwaltungs-Framework, das verwendet wird, um die inkrementelle Datenverarbeitung und die Entwicklung von Datenpipelines zu vereinfachen, indem Einfüge-, Aktualisierungs-, Upsert- und Löschfunktionen auf Datensatzebene bereitgestellt werden. Upsert bezieht sich auf die Möglichkeit, Datensätze in ein vorhandenes Dataset einzufügen, wenn sie noch nicht vorhanden sind, oder sie zu aktualisieren, falls sie schon vorhanden sind. Durch die effiziente Verwaltung der Datenaufstellung in Amazon S3 ermöglicht Hudi die Aufnahme und Aktualisierung von Daten in nahezu Echtzeit. Hudi verwaltet sorgfältig Metadaten der Aktionen, die für das Dataset ausgeführt werden, um leichter sicherzustellen zu können, dass die Aktionen atomar und konsistent sind.

Hudi ist in [Apache Spark](#), [Apache Hive](#) und [Presto](#) integriert. In den Amazon-EMR-Release-Versionen 6.1.0 und höher ist Hudi auch in [Trino \(PrestoSQL\)](#) integriert.

Ab Amazon-EMR-Version 5.28.0 und höher installiert EMR Hudi-Komponenten standardmäßig, wenn Spark, Hive Presto oder Flink installiert ist. Sie können Spark oder das DeltaStreamer Hudi-Hilfsprogramm verwenden, um Hudi-Datensätze zu erstellen oder zu aktualisieren. Sie können Hive, Spark, Presto oder Flink verwenden, um ein Hudi-Datensatz interaktiv abzufragen oder Datenverarbeitungs Pipelines mithilfe des inkrementellen Pull-Vorgangs zu erstellen. Inkrementeller Pull-Vorgang bezieht sich auf die Fähigkeit, nur die Daten zu extrahieren, die sich zwischen zwei Aktionen geändert haben.

Aufgrund dieser Features eignet sich Hudi für die folgenden Anwendungsfälle:

- Arbeiten mit Streaming-Daten von Sensoren und anderen IoT-Geräten (Internet of Things), die bestimmte Dateneinfüge- und Aktualisierungsereignisse erfordern.
- Einhaltung der Datenschutzvorschriften in Anwendungen, in denen Benutzer möglicherweise vergessen werden möchten oder ihre Einwilligung für die Verwendung ihrer Daten ändern möchten.
- Implementierung eines [CDC-Systems \(Change Data Capture\)](#), mit dem Sie Änderungen relativ zur Zeit auf ein Dataset anwenden können.

Die folgende Tabelle listet die Version von Hudi auf, die in der neuesten Version der Amazon-EMR-7.x-Serie enthalten ist, zusammen mit den Komponenten, die Amazon EMR mit Hudi installiert.

[Informationen zur Version der Komponenten, die in dieser Version mit Hudi installiert wurden, finden Sie unter Komponentenversionen von Version 7.1.0.](#)

## Hudi-Versionsinformationen für emr-7.1.0


Amazon-EMR-Version sbezeichnung	Hudi-Version	Mit Hudi installierte Komponenten
emr-7.1.0	Hudi 0.14.1-amzn-0	Not available.

Die folgende Tabelle listet die Version von Hudi auf, die in der neuesten Version der Amazon-EMR-6.x-Serie enthalten ist, zusammen mit den Komponenten, die Amazon EMR mit Hudi installiert.

Die Version der Komponenten, die mit Hudi in dieser Version installiert wurden, finden Sie unter [Komponentenversionen der Version 6.15.0](#).

## Hudi-Versionsinformationen für emr-6.15.0

Amazon-EMR-Version sbezeichnung	Hudi-Version	Mit Hudi installierte Komponenten
emr-6.15.0	Hudi 0.14.0-amzn-0	Not available.

 Note

Amazon-EMR-Version 6.8.0 enthält [Apache Hudi](#) 0.11.1. Amazon-EMR-6.8.0-Cluster sind jedoch auch mit der Open-Source-Version von Hudi 0.12.0 kompatibel. `hudi-spark3.3-bundle_2.12`

Die folgende Tabelle listet die Version von Hudi auf, die in der neuesten Version der Amazon-EMR-5.x-Serie enthalten ist, zusammen mit den Komponenten, die Amazon EMR mit Hudi installiert.

[Informationen zur Version der Komponenten, die in dieser Version mit Hudi installiert wurden, finden Sie unter Komponentenversionen von Version 5.36.2.](#)

## Hudi-Versionsinformationen für emr-5.36.2

Amazon-EMR-Version sbezeichnung	Hudi-Version	Mit Hudi installierte Komponenten
emr-5.36.2	Hudi 0.10.1-amzn-1	Not available.

## Themen

- [Funktionsweise von Hudi](#)
- [Überlegungen und Einschränkungen für die Verwendung von Hudi in Amazon EMR](#)
- [Erstellen eines Cluster mit installierter Hudi-Anwendung](#)
- [Arbeiten mit einem Hudi-Datensatz](#)
- [Die Hudi-CLI verwenden](#)
- [Hudi-Versionsverlauf](#)

## Funktionsweise von Hudi

Wenn Sie Hudi mit Amazon EMR verwenden, können Sie mit der Spark-Datenquellen-API oder dem DeltaStreamer Hudi-Hilfsprogramm Daten in den Datensatz schreiben. Hudi organisiert einen Datensatz in einer partitionierten Verzeichnisstruktur unter einer *basepath*, die einer herkömmlichen Hive-Tabelle ähnelt. Die Besonderheiten, wie die Daten in diesen Verzeichnissen als Dateien angelegt werden, hängen von dem von Ihnen ausgewählten Dataset-Typ ab. Sie können entweder Copy on Write (CoW, Kopieren beim Schreiben) oder Merge on Read (MoR, Beim Lesen zusammenführen) wählen.

Unabhängig vom Dataset-Typ wird jede Partition in einem Dataset eindeutig durch ihren *partitionpath* relativ zum *basepath* identifiziert. Innerhalb jeder Partition werden Datensätze in mehrere Datendateien verteilt. Weitere Informationen finden Sie unter [Dateimanagement](#) in der Apache Hudi-Dokumentation.

Jede Aktion in Hudi hat einen entsprechenden Commit, der durch einen monoton ansteigenden Zeitstempel gekennzeichnet ist, der als Instant bezeichnet wird. Hudi speichert eine Reihe aller am Datensatz ausgeführten Aktionen als Zeitleiste. Hudi nutzt die Timeline, um eine Snapshot-Isolation zwischen Lesern und Autoren zu ermöglichen und einen Rollback zu einem früheren Zeitpunkt zu ermöglichen. Weitere Informationen zu den von Hudi aufgezeichneten Aktionen und dem Status der Aktionen finden Sie unter [Timeline](#) in der Apache-Hudi-Dokumentation.

## Grundlegendes zu den Speichertypen von Datensätzen: „Beim Schreiben kopieren“ gegenüber „beim Lesen zusammenführen“

Wenn Sie ein Hudi-Datensatz erstellen, geben Sie an, dass das Datensatz entweder beim Schreiben kopiert oder beim Lesen zusammengeführt werden soll.

- Copy on Write (CoW, Beim Schreiben kopieren) – Daten werden in einem spaltenbasierten Format (Parquet) gespeichert, und jedes Update erstellt während eines Schreibvorgangs eine neue Version von Dateien. CoW ist der Standard Speichertyp.
- Merge on Read (MoR, Beim Lesen zusammenführen) – Daten werden mit einer Kombination aus spalten- (Parquet) und zeilenbasierten (Avro) Formaten gespeichert. Updates werden in zeilenbasierten Delta-Dateien protokolliert und nach Bedarf komprimiert, um neue Versionen der Spaltendateien zu erstellen.

Bei CoW-Datasets wird jedes Mal, wenn ein Datensatz aktualisiert wird, die Datei, die den Datensatz enthält, mit den aktualisierten Werten neu geschrieben. Bei einem MoR-Datensatz schreibt Hudi jedes Mal, wenn es eine Aktualisierung gibt, nur die Zeile für den geänderten Datensatz. MoR eignet sich besser für schreib- oder änderungsintensive Workloads mit weniger Lesevorgängen. CoW eignet sich besser für leseintensive Workloads für Daten, die sich seltener ändern.

Hudi bietet drei logische Ansichten für den Zugriff auf die Daten:

- Lese-optimierte Ansicht – Stellt den neuesten festgeschriebenen Datensatz aus CoW-Tabellen und den neuesten komprimierten Datensatz aus MoR-Tabellen bereit.
- Inkrementelle Ansicht – Stellt einen Änderungsstream zwischen zwei Aktionen aus einem CoW-Datensatz bereit, um nachgeschaltete Aufträge zu versorgen und Workflows zu extrahieren, transformieren, laden (ETL).
- Echtzeitansicht – Stellt die neuesten festgeschriebenen Daten aus einer MoR-Tabelle bereit, indem die spalten- und zeilenbasierten Dateien inline zusammengeführt werden.

Wenn Sie die leseoptimierte Ansicht abfragen, gibt die Abfrage alle komprimierten Daten zurück, enthält jedoch nicht die neuesten Delta-Commits. Das Abfragen dieser Daten bietet eine gute Leseleistung, aber die neuesten Daten werden nicht mehr angezeigt. Wenn Sie die Echtzeitansicht abfragen, führt Hudi die komprimierten Daten beim Lesen mit den Delta-Commits zusammen. Die neuesten Daten stehen für die Abfrage zur Verfügung, aber der Rechenaufwand für die Zusammenführung macht die Abfrage weniger performant. Die Möglichkeit, komprimierte Daten oder Echtzeitdaten abzufragen, ermöglicht es Ihnen, bei der Abfrage zwischen Leistung und Flexibilität zu wählen.

Weitere Informationen zu den Kompromissen zwischen Speichertypen finden Sie unter [Speichertypen und Ansichten](#) in der Apache-Hudi-Dokumentation.



Hudi erstellt zwei Tabellen im Hive-Metastore für MoR: eine Tabelle mit dem von Ihnen angegebenen Namen, bei der es sich um eine leseoptimierte Ansicht handelt, und eine Tabelle mit demselben Namen, an den `_rt` angehängt wird, bei der es sich um eine Echtzeitansicht handelt. Sie können beide Tabellen abfragen.

## Registrieren eines Hudi-Datensatzes bei Ihrem Metastore

Wenn Sie eine Hudi-Tabelle mit dem Hive-Metastore registrieren, können Sie Hudi-Tabellen wie jede andere Tabelle mit Hive, Spark SQL oder Presto abfragen. Darüber hinaus können Sie Hudi in AWS Glue integrieren, indem Sie Hive und Spark so konfigurieren, dass sie den AWS Glue-Datenkatalog als Metastore verwenden. Hudi registriert für MoR-Tabellen den Datensatz als zwei Tabellen im Metastore: eine Tabelle mit dem von Ihnen angegebenen Namen, bei der es sich um eine leseoptimierte Ansicht handelt, und eine Tabelle mit demselben Namen, an die `_rt` angehängt wird, bei dem es sich um eine Echtzeitansicht handelt.

Sie registrieren eine Hudi-Tabelle mit dem Hive-Metastore, wenn Sie mit Spark ein Hudi-Datensatz erstellen, indem Sie die Option `HIVE_SYNC_ENABLED_OPT_KEY` auf `"true"` einstellen und andere erforderliche Eigenschaften bereitstellen. Weitere Informationen finden Sie unter [Arbeiten mit einem Hudi-Datensatz](#). Darüber hinaus können Sie das Befehlszeilenprogramm `hive_sync_tool` verwenden, um einen Hudi-Datensatz separat als Tabelle in Ihrem Metastore zu registrieren.

## Überlegungen und Einschränkungen für die Verwendung von Hudi in Amazon EMR

- Datensatzschlüselfeld darf nicht null oder leer sein – Das Feld, das Sie als Datensatzschlüselfeld angeben, darf weder `null`-Werte noch leere Werte aufweisen.
- Das Schema wird standardmäßig beim Upsert und Insert aktualisiert — Hudi bietet eine Schnittstelle, die festlegt, wie `HoodieRecordPayload` die Eingabe `DataFrame` und der vorhandene Hudi-Datensatz zusammengeführt werden, um einen neuen, aktualisierten Datensatz zu erzeugen. Hudi bietet eine Standardimplementierung dieser Klasse `OverwriteWithLatestAvroPayload`, die bestehende Datensätze überschreibt und das in der Eingabe angegebene Schema aktualisiert. `DataFrame` Um diese Logik für die Implementierung von Merge- und Teilaktualisierungen anzupassen, können Sie mithilfe des Parameters `DataSourceWriteOptions.PAYLOAD_CLASS_OPT_KEY` eine Implementierung der `HoodieRecordPayload`-Schnittstelle bereitstellen.

- Löschen erfordert Schema – Beim Löschen müssen Sie den Datensatzschlüssel, den Partitionsschlüssel und die Schlüsselfelder für die Vorab-Kombination angeben. Andere Spalten können auf null gesetzt werden oder leer sein, aber das vollständige Schema ist erforderlich.
- Einschränkungen bei MoR-Tabellen – MoR-Tabellen unterstützen kein Savepointing. Sie können MoR-Tabellen mit der leseoptimierten Ansicht oder der Echtzeitansicht (*tableName\_rt*) von Spark SQL, Presto oder Hive abfragen. Wenn Sie die leseoptimierte Ansicht verwenden, werden nur Basisdateidaten und keine zusammengeführte Ansicht von Basis- und Protokolldaten angezeigt.
- Hive
  - Für die Registrierung von Tabellen im Hive-Metastore erwartet, dass der Hive Thrift-Server am Standardport 10000 ausgeführt wird. Wenn Sie diesen Port mit einem benutzerdefinierten Port überschreiben, übergeben Sie die Option HIVE\_URL\_OPT\_KEY wie im folgenden Beispiel gezeigt.

```
.option(DataSourceWriteOptions.HIVE_URL_OPT_KEY, "jdbc:hive2://localhost:override-port-number
```

- Der Datentyp `timestamp` in Spark ist in Hive als `long`-Datentyp und nicht als Hive-Typ `timestamp` registriert.
- Presto
  - Presto unterstützt nicht das Lesen von MoR-Echtzeittabellen.Funktionen in Hudi-Versionen unter 0.6.0.
  - Presto unterstützt nur Snapshot-Abfragen.
  - Damit Presto Hudi-Datensatz-Spalten korrekt interpretiert, stellen Sie den `hive.parquet_use_column_names`-Wert auf `true` ein.
    - Um den Wert für eine Sitzung festzulegen, führen Sie in der Presto-Shell den folgenden Befehl aus:

```
set session hive.parquet_use_column_names=true
```

- Um den Wert auf Clusterebene festzulegen, verwenden Sie die Konfigurationsklassifizierung `presto-connector-hive`, um `hive.parquet.use_column_names` auf `true` einzustellen, wie im folgenden Beispiel veranschaulicht. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

```
[
```

```
{
  "Classification": "presto-connector-hive",
  "Properties": {
    "hive.parquet.use-column-names": "true"
  }
}
```

- HBase-Index
- Die HBase-Version, die zum Erstellen von Hudi verwendet wurde, unterscheidet sich möglicherweise von der in den EMR-Versionshinweisen aufgeführten Version. Um die richtigen Abhängigkeiten für Ihre Spark-Sitzung abzurufen, führen Sie den folgenden Befehl aus.

```
spark-shell \  
--jars /usr/lib/spark/external/lib/spark-avro.jar,/usr/lib/hudi/cli/lib/*.jar \  
--conf "spark.serializer=org.apache.spark.serializer.KryoSerializer" \  
--conf "spark.sql.hive.convertMetastoreParquet=false"
```

## Erstellen eines Cluster mit installierter Hudi-Anwendung

Ab der Amazon-EMR-Version 5.28.0 installiert Amazon EMR standardmäßig Hudi-Komponenten, wenn Spark, Hive oder Presto installiert wird. Um Hudi in Amazon EMR zu verwenden, erstellen Sie einen Cluster mit den folgenden installierten Anwendungen:

- Hadoop
- Hive
- Spark
- Presto
- Flink

Sie können einen Cluster mithilfe der AWS Management Console AWS CLI, der oder der Amazon EMR-API erstellen.

## Um einen Cluster mit Hudi zu erstellen, verwenden Sie AWS Management Console

1. Navigieren Sie zur neuen Amazon-EMR-Konsole und wählen Sie in der Seitennavigation die Option Zur alten Konsole wechseln aus. Weitere Informationen darüber, was Sie erwartet, wenn Sie zur alten Konsole wechseln, finden Sie unter [Verwenden der alten Konsole](#).
2. Wählen Sie Create Cluster (Cluster erstellen) und Go to advanced options (Zu erweiterten Optionen) aus.
3. Wählen Sie unter „Software Configuration (Softwarekonfiguration)“ emr-5.28.0 oder höher für Release aus und wählen Sie Hadoop, Hive, Spark, Presto und Tez zusammen mit anderen Anwendungen aus, die Ihr Cluster benötigt.
4. Konfigurieren Sie nach Bedarf andere Optionen für Ihre Anwendung und wählen Sie dann Next (Weiter).
5. Konfigurieren Sie die Optionen für Hardware und General cluster settings (Allgemeine Clustereinstellungen) ganz nach Wunsch.
6. Für Security Options (Sicherheitsoptionen) empfehlen wir, ein EC2 key pair (EC2-Schlüsselpaar) auszuwählen, mit dem Sie mithilfe von SSH eine Verbindung mit der Befehlszeile des Master-Knotens herstellen können. Auf diese Weise können Sie die in diesem Handbuch beschriebenen Spark-Shell-Befehle, Hive-CLI-Befehle und Hudi-CLI-Befehle ausführen.
7. Wählen Sie weitere Sicherheitsoptionen wie gewünscht und wählen Sie anschließend Create cluster (Cluster erstellen) aus.

## Arbeiten mit einem Hudi-Datensatz

Hudi unterstützt das Einfügen, Aktualisieren und Löschen von Daten in Hudi-Datensätzen über Spark. Weitere Informationen finden Sie unter [Hudi-Tabellen schreiben](#) in der Apache-Hudi-Dokumentation.

In den folgenden Beispielen wird veranschaulicht, wie Sie die interaktive Spark-Shell starten und die Spark-Übermittlung und Amazon EMR Notebooks verwenden, um mit Hudi in Amazon EMR zu arbeiten. Sie können auch das DeltaStreamer Hudi-Hilfsprogramm oder andere Tools verwenden, um in einen Datensatz zu schreiben. Die Beispiele in diesem Abschnitt veranschaulichen das Arbeiten mit Datasets unter Verwendung der Spark-Shell, während eine Verbindung mit dem Master-Knoten mittels SSH als Standard-hadoop-Benutzer vorhanden ist.

## Starten Sie die Spark-Shell mit Amazon EMR 6.7 und höher

Wenn Sie `spark-shell`, `spark-submit` oder `spark-sql` mit Amazon EMR 6.7.0 oder höher ausführen, übergeben Sie die folgenden Befehle.

### Note

Amazon EMR 6.7.0 verwendet [Apache Hudi 0.11.0-amzn-0](#), das erhebliche Verbesserungen gegenüber früheren Hudi-Versionen enthält. Weitere Informationen finden Sie im [Apache-Hudi-0.11.0-Migrationshandbuch](#). Die Beispiele auf dieser Registerkarte spiegeln diese Änderungen wider.

So öffnen Sie die Spark-Shell auf dem Primärknoten

1. Verbinden Sie sich dem Primärknoten über SSH. Weitere Informationen finden Sie unter [Stellen Sie über SSH eine Verbindung zum Primärknoten her](#) im Verwaltungshandbuch für Amazon EMR.
2. Geben Sie den folgenden Befehl ein, um die Spark-Shell zu starten. *Um die PySpark Shell zu verwenden, ersetzen Sie `spark-shell` durch `pyspark`.*

```
spark-shell --jars /usr/lib/hudi/hudi-spark-bundle.jar \  
--conf "spark.serializer=org.apache.spark.serializer.KryoSerializer" \  
--conf \  
"spark.sql.catalog.spark_catalog=org.apache.spark.sql.hudi.catalog.HoodieCatalog" \  
\  
--conf "spark.sql.extensions=org.apache.spark.sql.hudi.HoodieSparkSessionExtension"
```

## Starten Sie die Spark-Shell mit Amazon EMR 6.6 und früher

Wenn Sie `spark-shell`, `spark-submit`, oder `spark-sql` mit Amazon EMR 6.6.x oder früher ausführen, übergeben Sie die folgenden Befehle.

### Note

- Amazon EMR 6.2 und 5.31 und höher (Hudi 0.6.x und höher) können `spark-avro.jar` in der Konfiguration weglassen.

- Amazon EMR 6.5 und 5.35 und höher (Hudi 0.9.x und höher) können `spark.sql.hive.convertMetastoreParquet=false` in der Konfiguration weglassen.
- Amazon EMR 6.6 und 5.36 und höher (Hudi 0.10.x und höher) müssen die `HoodieSparkSessionExtension` Konfiguration enthalten, wie sie im [Spark-Leitfaden zu Version: 0.10.0](#) beschrieben ist:

```
--conf  
"spark.sql.extensions=org.apache.spark.sql.hudi.HoodieSparkSessionExtension"  
\
```

So öffnen Sie die Spark-Shell auf dem Primärknoten

1. Verbinden Sie sich dem Primärknoten über SSH. Weitere Informationen finden Sie unter [Stellen Sie über SSH eine Verbindung zum Primärknoten her](#) im Verwaltungshandbuch für Amazon EMR.
2. Geben Sie den folgenden Befehl ein, um die Spark-Shell zu starten. *Um die Shell zu verwenden, ersetzen Sie `PySpark spark-shell` durch `pyspark`.*

```
spark-shell \  
--conf "spark.serializer=org.apache.spark.serializer.KryoSerializer" \  
--conf "spark.sql.hive.convertMetastoreParquet=false" \  
--jars /usr/lib/hudi/hudi-spark-bundle.jar,/usr/lib/spark/external/lib/spark-  
avro.jar
```

## Verwenden Sie Hudi mit Amazon EMR Notebooks, die Amazon EMR 6.7 und höher verwenden

Um Hudi mit Amazon EMR Notebooks verwenden zu können, müssen Sie zuerst die Hudi Jar-Dateien aus dem lokalen Dateisystem auf HDFS auf dem Hauptknoten des Notebook-Clusters kopieren. Anschließend verwenden Sie den Notebook-Editor, um Ihr EMR Notebook für die Verwendung von Hudi zu konfigurieren.

## Wie Sie Hudi mit Amazon EMR Notebooks verwenden

1. Erstellen und starten Sie einen Cluster für Amazon EMR Notebooks. Weitere Informationen finden Sie unter [Erstellen von Amazon-EMR-Clustern für Notebooks](#) im Verwaltungshandbuch für Amazon EMR.
2. Verbinden Sie sich über SSH mit dem Master-Knoten des Clusters und kopieren Sie dann die Jar-Dateien aus dem lokalen Dateisystem auf HDFS, wie in den folgenden Beispielen veranschaulicht. In diesem Beispiel erstellen wir ein Verzeichnis in HDFS zur Veranschaulichung der Dateiverwaltung. Sie können Ihr eigenes Ziel in HDFS wählen, falls gewünscht.

```
hdfs dfs -mkdir -p /apps/hudi/lib
```

```
hdfs dfs -copyFromLocal /usr/lib/hudi/hudi-spark-bundle.jar /apps/hudi/lib/hudi-spark-bundle.jar
```

3. Öffnen Sie den Notebook-Editor, geben Sie den Code aus dem folgenden Beispiel ein und führen Sie ihn aus.

```
%%configure
{ "conf": {
    "spark.jars": "hdfs:///apps/hudi/lib/hudi-spark-bundle.jar",
    "spark.serializer": "org.apache.spark.serializer.KryoSerializer",
    "spark.sql.catalog.spark_catalog":
    "org.apache.spark.sql.hudi.catalog.HoodieCatalog",

    "spark.sql.extensions": "org.apache.spark.sql.hudi.HoodieSparkSessionExtension"
  }}
```

## Verwenden Sie Hudi mit Amazon EMR Notebooks, die Amazon EMR 6.6 und früher verwenden

Um Hudi mit Amazon EMR Notebooks verwenden zu können, müssen Sie zuerst die Hudi Jar-Dateien aus dem lokalen Dateisystem auf HDFS auf dem Hauptknoten des Notebook-Clusters kopieren. Anschließend verwenden Sie den Notebook-Editor, um Ihr EMR Notebook für die Verwendung von Hudi zu konfigurieren.

## Wie Sie Hudi mit Amazon EMR Notebooks verwenden

1. Erstellen und starten Sie einen Cluster für Amazon EMR Notebooks. Weitere Informationen finden Sie unter [Erstellen von Amazon-EMR-Clustern für Notebooks](#) im Verwaltungshandbuch für Amazon EMR.
2. Verbinden Sie sich über SSH mit dem Master-Knoten des Clusters und kopieren Sie dann die Jar-Dateien aus dem lokalen Dateisystem auf HDFS, wie in den folgenden Beispielen veranschaulicht. In diesem Beispiel erstellen wir ein Verzeichnis in HDFS zur Veranschaulichung der Dateiverwaltung. Sie können Ihr eigenes Ziel in HDFS wählen, falls gewünscht.

```
hdfs dfs -mkdir -p /apps/hudi/lib
```

```
hdfs dfs -copyFromLocal /usr/lib/hudi/hudi-spark-bundle.jar /apps/hudi/lib/hudi-spark-bundle.jar
```

```
hdfs dfs -copyFromLocal /usr/lib/spark/external/lib/spark-avro.jar /apps/hudi/lib/spark-avro.jar
```

3. Öffnen Sie den Notebook-Editor, geben Sie den Code aus dem folgenden Beispiel ein und führen Sie ihn aus.

```
{ "conf": {  
    "spark.jars": "hdfs:///apps/hudi/lib/hudi-spark-bundle.jar,hdfs:///apps/hudi/lib/spark-avro.jar",  
    "spark.serializer": "org.apache.spark.serializer.KryoSerializer",  
    "spark.sql.hive.convertMetastoreParquet": "false"  
}}
```

## Initialisieren Sie eine Spark-Sitzung für Hudi

Wenn Sie Scala verwenden, müssen Sie die folgenden Klassen in Ihre Spark-Sitzung importieren. Dies muss einmal pro Spark-Sitzung erfolgen.

```
import org.apache.spark.sql.SaveMode  
import org.apache.spark.sql.functions._  
import org.apache.hudi.DataSourceWriteOptions  
import org.apache.hudi.DataSourceReadOptions  
import org.apache.hudi.config.HoodieWriteConfig
```



```
import org.apache.hudi.hive.MultiPartKeyValueExtractor
import org.apache.hudi.hive.HiveSyncConfig
import org.apache.hudi.sync.common.HoodieSyncConfig
```

## Schreiben Sie in einen Hudi-Datensatz

Die folgenden Beispiele zeigen, wie ein Hudi-Datensatz erstellt DataFrame und als Hudi-Datensatz geschrieben wird.

### Note

Um Codebeispiele in die Spark-Shell einzufügen, geben Sie an der Eingabeaufforderung **:paste** ein, fügen das Beispiel ein und drücken dann **CTRL + D**.

Jedes Mal, wenn Sie einen DataFrame in einen Hudi-Datensatz schreiben, müssen Sie Folgendes angeben. `DataSourceWriteOptions` Viele dieser Optionen sind unter den Schreiboperationen wahrscheinlich identisch. Im folgenden Beispiel werden allgemeine Optionen unter Verwendung der Variablen `hoodieOptions` angegeben, die von nachfolgenden Beispielen verwendet wird.

Schreiben Sie mit Scala mit Amazon EMR 6.7 und höher

### Note

Amazon EMR 6.7.0 verwendet [Apache Hudi 0.11.0-amzn-0](#), das erhebliche Verbesserungen gegenüber früheren Hudi-Versionen enthält. Weitere Informationen finden Sie im [Apache-Hudi-0.11.0-Migrationshandbuch](#). Die Beispiele auf dieser Registerkarte spiegeln diese Änderungen wider.

```
// Create a DataFrame
val inputDF = Seq(
  ("100", "2015-01-01", "2015-01-01T13:51:39.340396Z"),
  ("101", "2015-01-01", "2015-01-01T12:14:58.597216Z"),
  ("102", "2015-01-01", "2015-01-01T13:51:40.417052Z"),
  ("103", "2015-01-01", "2015-01-01T13:51:40.519832Z"),
  ("104", "2015-01-02", "2015-01-01T12:15:00.512679Z"),
  ("105", "2015-01-02", "2015-01-01T13:51:42.248818Z")
).toDF("id", "creation_date", "last_update_time")
```

```
//Specify common DataSourceWriteOptions in the single hudiOptions variable
val hudiOptions = Map[String,String](
  HoodieWriteConfig.TBL_NAME.key -> "tableName",
  DataSourceWriteOptions.TABLE_TYPE.key -> "COPY_ON_WRITE",
  DataSourceWriteOptions.RECORDKEY_FIELD_OPT_KEY -> "id",
  DataSourceWriteOptions.PARTITIONPATH_FIELD_OPT_KEY -> "creation_date",
  DataSourceWriteOptions.PRECOMBINE_FIELD_OPT_KEY -> "last_update_time",
  DataSourceWriteOptions.HIVE_SYNC_ENABLED_OPT_KEY -> "true",
  DataSourceWriteOptions.HIVE_TABLE_OPT_KEY -> "tableName",
  DataSourceWriteOptions.HIVE_PARTITION_FIELDS_OPT_KEY -> "creation_date",
  HoodieSyncConfig.META_SYNC_PARTITION_EXTRACTOR_CLASS.key ->
"org.apache.hudi.hive.MultiPartKeysValueExtractor",
  HoodieSyncConfig.META_SYNC_ENABLED.key -> "true",
  HiveSyncConfig.HIVE_SYNC_MODE.key -> "hms",
  HoodieSyncConfig.META_SYNC_TABLE_NAME.key -> "tableName",
  HoodieSyncConfig.META_SYNC_PARTITION_FIELDS.key -> "creation_date"
)

// Write the DataFrame as a Hudi dataset
(inputDF.write
  .format("hudi")
  .options(hudiOptions)
  .option(DataSourceWriteOptions.OPERATION_OPT_KEY,"insert")
  .mode(SaveMode.Overwrite)
  .save("s3://DOC-EXAMPLE-BUCKET/myhudidataset/"))
```

## Schreiben Sie mit Scala mit Amazon EMR 6.6 und früher

```
// Create a DataFrame
val inputDF = Seq(
  ("100", "2015-01-01", "2015-01-01T13:51:39.340396Z"),
  ("101", "2015-01-01", "2015-01-01T12:14:58.597216Z"),
  ("102", "2015-01-01", "2015-01-01T13:51:40.417052Z"),
  ("103", "2015-01-01", "2015-01-01T13:51:40.519832Z"),
  ("104", "2015-01-02", "2015-01-01T12:15:00.512679Z"),
  ("105", "2015-01-02", "2015-01-01T13:51:42.248818Z")
).toDF("id", "creation_date", "last_update_time")

//Specify common DataSourceWriteOptions in the single hudiOptions variable
val hudiOptions = Map[String,String](
  HoodieWriteConfig.TABLE_NAME -> "tableName",
  DataSourceWriteOptions.TABLE_TYPE_OPT_KEY -> "COPY_ON_WRITE",
```

```

DataSourceWriteOptions.RECORDKEY_FIELD_OPT_KEY -> "id",
DataSourceWriteOptions.PARTITIONPATH_FIELD_OPT_KEY -> "creation_date",
DataSourceWriteOptions.PRECOMBINE_FIELD_OPT_KEY -> "last_update_time",
DataSourceWriteOptions.HIVE_SYNC_ENABLED_OPT_KEY -> "true",
DataSourceWriteOptions.HIVE_TABLE_OPT_KEY -> "tableName",
DataSourceWriteOptions.HIVE_PARTITION_FIELDS_OPT_KEY -> "creation_date",
DataSourceWriteOptions.HIVE_PARTITION_EXTRACTOR_CLASS_OPT_KEY ->
classOf[MultiPartKeysValueExtractor].getName
)

// Write the DataFrame as a Hudi dataset
(inputDF.write
  .format("org.apache.hudi")
  .option(DataSourceWriteOptions.OPERATION_OPT_KEY,
DataSourceWriteOptions.INSERT_OPERATION_OPT_VAL)
  .options(hudiOptions)
  .mode(SaveMode.Overwrite)
  .save("s3://DOC-EXAMPLE-BUCKET/myhudidataset/"))

```

## Schreiben Sie mit PySpark

```

# Create a DataFrame
inputDF = spark.createDataFrame(
  [
    ("100", "2015-01-01", "2015-01-01T13:51:39.340396Z"),
    ("101", "2015-01-01", "2015-01-01T12:14:58.597216Z"),
    ("102", "2015-01-01", "2015-01-01T13:51:40.417052Z"),
    ("103", "2015-01-01", "2015-01-01T13:51:40.519832Z"),
    ("104", "2015-01-02", "2015-01-01T12:15:00.512679Z"),
    ("105", "2015-01-02", "2015-01-01T13:51:42.248818Z"),
  ],
  ["id", "creation_date", "last_update_time"]
)

# Specify common DataSourceWriteOptions in the single hudiOptions variable
hudiOptions = {
'hoodie.table.name': 'tableName',
'hoodie.datasource.write.recordkey.field': 'id',
'hoodie.datasource.write.partitionpath.field': 'creation_date',
'hoodie.datasource.write.precombine.field': 'last_update_time',
'hoodie.datasource.hive_sync.enable': 'true',
'hoodie.datasource.hive_sync.table': 'tableName',
'hoodie.datasource.hive_sync.partition_fields': 'creation_date',

```

```
'hoodie.datasource.hive_sync.partition_extractor_class':
  'org.apache.hudi.hive.MultiPartKeysValueExtractor'
}

# Write a DataFrame as a Hudi dataset
inputDF.write \
  .format('org.apache.hudi') \
  .option('hoodie.datasource.write.operation', 'insert') \
  .options(**hudiOptions) \
  .mode('overwrite') \
  .save('s3://DOC-EXAMPLE-BUCKET/myhudidataset/')
```

### Note

Möglicherweise sehen Sie in Codebeispielen und Benachrichtigungen „hoodie“ anstelle von Hudi. In der Hudi-Codebasis wird häufig die alte Schreibweise „hoodie“ verwendet.

## DataSourceWriteOptions Referenz für Hudi

Option	Beschreibung
TABLE_NAME	Der Tabellename, unter dem das Dataset registriert werden soll.
TABLE_TYPE_OPT_KEY	Optional. Gibt an, ob das Dataset als "COPY_ON_WRITE" oder "MERGE_ON_READ" erstellt wird. Der Standardwert ist "COPY_ON_WRITE" .
RECORDKEY_FIELD_OPT_KEY	Das Datensatzschlüselfeld, dessen Wert als die <code>recordKey</code> -Komponente von <code>HoodieKey</code> verwendet wird. Der tatsächliche Wert wird durch Aufrufen von <code>.toString()</code> für den Feldwert abgerufen. Verschachtelte Felder können mithilfe der Punktnotation angegeben werden, z. B <code>a.b.c</code> .
PARTITIONPATH_FIELD_OPT_KEY	Das Partitionsfeld, dessen Wert als die Komponente <code>partitionPath</code> von

Option	Beschreibung
PRECOMBINE_FIELD_OPT_KEY	<p>HoodieKey verwendet wird. Der tatsächliche Wert wird durch Aufrufen von <code>.toString()</code> für den Feldwert abgerufen.</p> <p>Das Feld, das in der Vorab-Kombination vor dem tatsächlichen Schreiben verwendet wird. Wenn zwei Datensätze denselben Schlüsselwert haben, wählt Hudi den Datensatz mit dem größten Wert für das Vorab-Kombinationsfeld wie von <code>Object.compareTo(..)</code> festgelegt aus.</p>

Die folgenden Optionen sind nur erforderlich, um die Hudi-Datensatz-Tabelle in Ihrem Metastore zu registrieren. Wenn Sie Ihr Hudi-Datensatz nicht als Tabelle im Hive-Metastore registrieren, sind diese Optionen nicht erforderlich.

#### DataSourceWriteOptions Referenz für Hive

Option	Beschreibung
HIVE_DATABASE_OPT_KEY	Die Hive-Datenbank, mit der synchronisiert werden soll. Der Standardwert ist "default".
HIVE_PARTITION_EXTRACTOR_CLASS_OPT_KEY	Die Klasse, mit der Partitionsfeldwerte in Hive-Partitionsspalten extrahiert werden.
HIVE_PARTITION_FIELDS_OPT_KEY	Das Feld im Dataset, anhand dessen Hive-Partitionsspalten bestimmt werden sollen.
HIVE_SYNC_ENABLED_OPT_KEY	Wenn diese Option auf "true" eingestellt ist, wird das Dataset beim Apache-Hive-Metastore registriert. Der Standardwert ist "false".

Option	Beschreibung
HIVE_TABLE_OPT_KEY	Erforderlich Der Name der Tabelle in Hive, mit der synchronisiert werden soll. z. B. "my_hudi_table_cow" .
HIVE_USER_OPT_KEY	Optional. Der Hive-Benutzername, der bei der Synchronisierung verwendet werden soll. z. B. "hadoop".
HIVE_PASS_OPT_KEY	Optional. Das von HIVE_USER_OPT_KEY angegebene Hive-Passwort für den Benutzer.
HIVE_URL_OPT_KEY	Die Hive-Metastore-URL.

## Upsert Daten

Das folgende Beispiel zeigt, wie Daten durch das Schreiben von A verändert werden. DataFrame  
Im Gegensatz zum vorherigen Einfügebeispiel wird der Wert OPERATION\_OPT\_KEY auf UPSERT\_OPERATION\_OPT\_VAL eingestellt. Darüber hinaus wird mit `.mode(SaveMode.Append)` angegeben, dass der Datensatz angehängt werden soll.

Upsert mit Scala mit Amazon EMR 6.7 und höher

### Note

Amazon EMR 6.7.0 verwendet [Apache Hudi 0.11.0-amzn-0](#), das erhebliche Verbesserungen gegenüber früheren Hudi-Versionen enthält. Weitere Informationen finden Sie im [Apache-Hudi-0.11.0-Migrationshandbuch](#). Die Beispiele auf dieser Registerkarte spiegeln diese Änderungen wider.

```
// Create a new DataFrame from the first row of inputDF with a different creation_date value
val updatedDF = inputDF.limit(1).withColumn("creation_date", lit("new_value"))

(updatedDF.write
  .format("hudi"))
```

```
.options(hudiOptions)
.option(DataSourceWriteOptions.OPERATION_OPT_KEY, "upsert")
.mode(SaveMode.Append)
.save("s3://DOC-EXAMPLE-BUCKET/myhuidataset/"))
```

## Upsert mit Scala mit Amazon EMR 6.6 und früher

```
// Create a new DataFrame from the first row of inputDF with a different creation_date
value
val updateDF = inputDF.limit(1).withColumn("creation_date", lit("new_value"))

(updateDF.write
  .format("org.apache.hudi")
  .option(DataSourceWriteOptions.OPERATION_OPT_KEY,
DataSourceWriteOptions.UPSERT_OPERATION_OPT_VAL)
  .options(hudiOptions)
  .mode(SaveMode.Append)
  .save("s3://DOC-EXAMPLE-BUCKET/myhuidataset/"))
```

## Upsert mit PySpark

```
from pyspark.sql.functions import lit

# Create a new DataFrame from the first row of inputDF with a different creation_date
value
updateDF = inputDF.limit(1).withColumn('creation_date', lit('new_value'))

updateDF.write \
  .format('org.apache.hudi') \
  .option('hoodie.datasource.write.operation', 'upsert') \
  .options(**hudiOptions) \
  .mode('append') \
  .save('s3://DOC-EXAMPLE-BUCKET/myhuidataset/')
```

## Einen Datensatz löschen

Um einen Datensatz dauerhaft zu löschen, können Sie eine leere Datenlast einfügen. In diesem Fall gibt die Option `PAYLOAD_CLASS_OPT_KEY` die Klasse `EmptyHoodieRecordPayload` an. Im Beispiel wird derselbe `DataFrame`, verwendet `updateDF`, der im Upsert-Beispiel verwendet wurde, um denselben Datensatz anzugeben.

## Löschen mit Scala mit Amazon EMR 6.7 und höher

### Note

Amazon EMR 6.7.0 verwendet [Apache Hudi](#) 0.11.0-amzn-0, das erhebliche Verbesserungen gegenüber früheren Hudi-Versionen enthält. Weitere Informationen finden Sie im [Apache-Hudi-0.11.0-Migrationshandbuch](#). Die Beispiele auf dieser Registerkarte spiegeln diese Änderungen wider.

```
(updateDF.write
  .format("hudi")
  .options(hudiOptions)
  .option(DataSourceWriteOptions.OPERATION_OPT_KEY, "delete")
  .mode(SaveMode.Append)
  .save("s3://DOC-EXAMPLE-BUCKET/myhuidataset/"))
```

## Löschen mit Scala mit Amazon EMR 6.6 und früher

```
(updateDF.write
  .format("org.apache.hudi")
  .option(DataSourceWriteOptions.OPERATION_OPT_KEY,
DataSourceWriteOptions.UPSERT_OPERATION_OPT_VAL)
  .option(DataSourceWriteOptions.PAYLOAD_CLASS_OPT_KEY,
"org.apache.hudi.common.model.EmptyHoodieRecordPayload")
  .mode(SaveMode.Append)
  .save("s3://DOC-EXAMPLE-BUCKET/myhuidataset/"))
```

## Löschen mit PySpark

```
updateDF.write \
  .format('org.apache.hudi') \
  .option('hoodie.datasource.write.operation', 'upsert') \
  .option('hoodie.datasource.write.payload.class',
'org.apache.hudi.common.model.EmptyHoodieRecordPayload') \
  .options(**hudiOptions) \
  .mode('append') \
  .save('s3://DOC-EXAMPLE-BUCKET/myhuidataset/')
```



Sie können Daten auch dauerhaft löschen, indem Sie `OPERATION_OPT_KEY` auf `DELETE_OPERATION_OPT_VAL` einstellen, um alle Datensätze in dem von Ihnen eingereichten Datensatz zu entfernen. Anweisungen zur Durchführung von „weichen Löschungen“ und weitere Informationen zum Löschen von Daten, die in Hudi-Tabellen gespeichert sind, finden Sie unter [Löschen](#) in der Apache-Hudi-Dokumentation.

## Aus einem Hudi-Datensatz lesen

Um Daten zum aktuellen Zeitpunkt abzurufen, führt Hudi standardmäßig Snapshot-Abfragen durch. Im Folgenden finden Sie ein Beispiel für die Abfrage des in S3 geschriebenen Datensatzes in [Schreiben Sie in einen Hudi-Datensatz](#). Ersetzen Sie `s3://DOC-EXAMPLE-BUCKET/myhuidataset` durch Ihren Tabellenpfad und fügen Sie Platzhalter-Sternchen für jede Partitionsebene sowie ein zusätzliches Sternchen hinzu. In diesem Beispiel gibt es eine Partitionsebene, daher haben wir zwei Platzhaltersymbole hinzugefügt.

Lesen Sie mit Scala mit Amazon EMR 6.7 und höher

### Note

Amazon EMR 6.7.0 verwendet [Apache Hudi 0.11.0-amzn-0](#), das erhebliche Verbesserungen gegenüber früheren Hudi-Versionen enthält. Weitere Informationen finden Sie im [Apache-Hudi-0.11.0-Migrationshandbuch](#). Die Beispiele auf dieser Registerkarte spiegeln diese Änderungen wider.

```
val snapshotQueryDF = spark.read
  .format("hudi")
  .load(s3://DOC-EXAMPLE-BUCKET/myhuidataset)
  .show()
```

Lesen mit Scala mit Amazon EMR 6.6 und früher

```
(val snapshotQueryDF = spark.read
  .format("org.apache.hudi")
  .load("s3://DOC-EXAMPLE-BUCKET/myhuidataset" + "/*/*"))

snapshotQueryDF.show()
```

## Lesen mit PySpark

```
snapshotQueryDF = spark.read \  
    .format('org.apache.hudi') \  
    .load('s3://DOC-EXAMPLE-BUCKET/myhuidataset' + '/*/*')  
  
snapshotQueryDF.show()
```

## Inkrementelle Abfragen

Sie können mit Hudi auch inkrementelle Abfragen durchführen, um einen Stream von Datensätzen abzurufen, die sich seit einem bestimmten Commit-Zeitstempel geändert haben. Setzen Sie dazu das Feld `QUERY_TYPE_OPT_KEY` auf `QUERY_TYPE_INCREMENTAL_OPT_VAL`. Fügen Sie dann einen Wert für `BEGIN_INSTANTTIME_OPT_KEY` hinzu, um alle Datensätze abzurufen, die seit dem angegebenen Zeitpunkt geschrieben wurden. Inkrementelle Abfragen sind in der Regel zehnmal effizienter als ihre Gegenstücke im Batch-Modus, da sie nur geänderte Datensätze verarbeiten.

Wenn Sie inkrementelle Abfragen ausführen, verwenden Sie den Pfad der Stammtabelle (Basistabelle) ohne die für Snapshot-Abfragen verwendeten Platzhaltersterchen.

### Note

Presto unterstützt keine inkrementellen Abfragen.

## Inkrementelle Abfragen mit Scala

```
(val incQueryDF = spark.read  
    .format("org.apache.hudi")  
    .option(DataSourceReadOptions.QUERY_TYPE_OPT_KEY,  
    DataSourceReadOptions.QUERY_TYPE_INCREMENTAL_OPT_VAL)  
    .option(DataSourceReadOptions.BEGIN_INSTANTTIME_OPT_KEY, <beginInstantTime>)  
    .load("s3://DOC-EXAMPLE-BUCKET/myhuidataset" ))  
  
incQueryDF.show()
```

## Inkrementelle Abfragen mit PySpark

```
readOptions = {  
    'hoodie.datasource.query.type': 'incremental',
```

```
'hoodie.datasource.read.begin.instanttime': <beginInstantTime>,  
}  
  
incQueryDF = spark.read \  
  .format('org.apache.hudi') \  
  .options(**readOptions) \  
  .load('s3://DOC-EXAMPLE-BUCKET/myhudidataset')  
  
incQueryDF.show()
```

Weitere Informationen zum Lesen von Hudi-Datensätzen finden Sie unter [Abfragen von Hudi-Tabellen](#) in der Apache-Hudi-Dokumentation.

## Die Hudi-CLI verwenden

Sie können mit der Hudi-CLI Hudi-Datensätze verwalten, um Informationen zu Commits, zum Dateisystem, zu Statistiken und mehr anzuzeigen. Sie können mit der CLI auch Komprimierungen manuell durchführen, Komprimierungen planen oder geplante Komprimierungen abrechnen. Weitere Informationen finden Sie unter [Dateimanagement](#) in der Apache-Hudi-Dokumentation.

So starten Sie die Hudi-CLI und stellen Sie eine Verbindung mit einem Datensatz her

1. Verbinden mit dem Master-Knoten über SSH. Weitere Informationen finden Sie unter [Mit SSH eine Verbindung zum Hauptknoten herstellen](#) im Verwaltungshandbuch für Amazon EMR.
2. Geben Sie in die Befehlszeile `/usr/lib/hudi/cli/bin/hudi-cli.sh` ein.

Die Eingabeaufforderung ändert sich in `hudi->`.

3. Geben Sie den folgenden Code ein, um eine Verbindung mit einem Dataset herzustellen. Ersetzen Sie `s3://DOC-EXAMPLE-BUCKET/myhudidataset` durch den Pfad zum Datensatz, mit dem Sie arbeiten möchten. Der von uns verwendete Wert entspricht dem Wert, der in früheren Beispielen festgelegt wurde.

```
connect --path s3://DOC-EXAMPLE-BUCKET/myhudidataset
```

Die Eingabeaufforderung ändert sich, um das Dataset einzuschließen, mit dem Sie verbunden sind, wie im folgenden Beispiel gezeigt.

```
hudi:myhudidataset->
```

## Hudi-Versionsverlauf

In der folgenden Tabelle sind die Version von Hudi aufgeführt, die in jeder Release-Version von Amazon EMR enthalten ist, zusammen mit den Komponenten, die mit der Anwendung installiert wurden. Informationen zu den Komponentenversionen in den einzelnen Versionen finden Sie im Abschnitt [Komponentenversion für Ihre Version in Amazon-EMR-7.x-Versionen](#), [Amazon-EMR-6.x-Versionen](#) oder [Amazon-EMR-5.x-Versionen](#).

### Versionsinformationen von Hudi

Amazon-EMR-Version sbezeichnung	Hudi-Version	Mit Hudi installierte Komponenten
emr-5.36.2	0.10.1-amzn-1	Not available.
emr-7.1.0	0.14.1-amzn-0	Not available.
emr-7.0.0	0.14.0-amzn-1	Not available.
emr-6.15.0	0.14.0-amzn-0	Not available.
emr-6.14.0	0.13.1-amzn-2	Not available.
emr-6.13.0	0.13.1-amzn-1	Not available.
emr-6.12.0	0.13.1-amzn-0	Not available.
emr-6.11.1	0.13.0-amzn-0	Not available.
emr-6.11.0	0.13.0-amzn-0	Not available.
emr-6.10.1	0.12.2-amzn-0	Not available.
emr-6.10.0	0.12.2-amzn-0	Not available.
emr-6.9.1	0.12.1-amzn-0	Not available.
emr-6.9.0	0.12.1-amzn-0	Not available.
emr-6.8.1	0.11.1-amzn-0	Not available.
emr-6.8.0	0.11.1-amzn-0	Not available.

Amazon-EMR-Version sbezeichnung	Hudi-Version	Mit Hudi installierte Komponenten
emr-6.7.0	0.11.0-amzn-0	Not available.
emr-5.36.1	0.10.1-amzn-1	Not available.
emr-5.36.0	0.10.1-amzn-1	Not available.
emr-6.6.0	0.10.1-amzn-0	Not available.
emr-5.35.0	0.9.0-amzn-2	Not available.
emr-6.5.0	0.9.0-amzn-1	Not available.
emr-6.4.0	0.8.0-amzn-0	Not available.
emr-6.3.1	0.7.0-amzn-0	Not available.
emr-6.3.0	0.7.0-amzn-0	Not available.
emr-6.2.1	0.6.0-amzn-1	Not available.
emr-6.2.0	0.6.0-amzn-1	Not available.
emr-6.1.1	0.5.2-incubating-amzn-2	Not available.
emr-6.1.0	0.5.2-incubating-amzn-2	Not available.
emr-6.0.1	0.5.0-incubating-amzn-1	Not available.
emr-6.0.0	0.5.0-incubating-amzn-1	Not available.
emr-5.34.0	0.9.0-amzn-0	Not available.
emr-5.33.1	0.7.0-amzn-1	Not available.
emr-5.33.0	0.7.0-amzn-1	Not available.
emr-5.32.1	0.6.0-amzn-0	Not available.
emr-5.32.0	0.6.0-amzn-0	Not available.

Amazon-EMR-Version sbezeichnung	Hudi-Version	Mit Hudi installierte Komponenten
emr-5.31.1	0.6.0-amzn-0	Not available.
emr-5.31.0	0.6.0-amzn-0	Not available.
emr-5.30.2	0.5.2-incubating	Not available.
emr-5.30.1	0.5.2-incubating	Not available.
emr-5.30.0	0.5.2-incubating	Not available.
emr-5.29.0	0.5.0-incubating	Not available.
emr-5.28.1	0.5.0-incubating	Not available.
emr-5.28.0	0.5.0-incubating	Not available.

# Hue

Hue (Hadoop User Experience) ist eine webbasierte, grafische Open-Source-Benutzeroberfläche für die Verwendung mit Amazon EMR und Apache Hadoop. Hue gruppiert verschiedene Projekte aus dem Hadoop-Ökosystem in einer konfigurierbaren Schnittstelle. Amazon EMR hat auch Anpassungen speziell für Hue in Amazon EMR hinzugefügt. Hue fungiert als Front-End für Anwendungen, die auf Ihrem Cluster ausgeführt werden, und ermöglicht Ihnen die Interaktion mit Anwendungen über eine bekanntere oder benutzerfreundlichere Benutzeroberfläche. Durch die Anwendungen in Hue, wie beispielsweise Hive- und Pig-Editoren, entfällt das Anmelden im Cluster, um Skripts interaktiv an der Shell der jeweiligen Anwendung auszuführen. Nachdem ein Cluster gestartet wurde, können Sie mit Hue oder einer ähnlichen Schnittstelle vollständig mit Anwendungen interagieren. Weitere Informationen zu Hue finden Sie unter <http://gethue.com>.

Hue ist standardmäßig installiert, wenn Sie Ihren Cluster über die Amazon-EMR-Konsole starten. Sie können mit Erweiterte Optionen in der Amazon-EMR-Konsole wählen, dass Hue nicht installiert wird, wenn Sie einen Cluster starten, oder indem Sie explizit die Option `--applications` festlegen und Hue auslassen, wenn Sie `create-cluster` über die AWS CLI verwenden.

## Themen

- [Die Hue Versionsinformationen](#)
- [Unterstützte und nicht unterstützte Features von Hue in Amazon EMR](#)
- [Überlegungen](#)
- [Verbindung zur Hue-Webbenutzeroberfläche herstellen](#)
- [Verwenden von Hue mit einer Remote-Datenbank in Amazon RDS](#)
- [Erweiterte Konfigurationen für Hue](#)
- [Hue-Versionsverlauf](#)

## Die Hue Versionsinformationen

### Hue-Version für 7.1.0

Die folgende Tabelle listet die Version von Hue auf, die in der neuesten Version der Amazon-EMR-7.x-Serie enthalten ist, zusammen mit den Komponenten, die Amazon EMR mit Hue installiert.

Die Version der Komponenten, die in dieser Version mit Hue installiert wurden, finden Sie unter [Komponentenversionen von Version 7.1.0](#).

## Hue-Versionsinformationen für emr-7.1.0

Amazon-EMR-Version sbezeichnung	Hue-Version	Mit Hue installierte Komponent en
emr-7.1.0	Hue 4.11.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

## Hue-Version für 6.15.0

Die folgende Tabelle listet die Version von Hue auf, die in der neuesten Version der Amazon-EMR-6.x-Serie enthalten ist, zusammen mit den Komponenten, die Amazon EMR mit Hue installiert.

Die Version der Komponenten, die mit Hue in dieser Version installiert wurden, finden Sie unter [Komponentenversionen der Version 6.15.0](#).

## Hue-Versionsinformationen für emr-6.15.0

Amazon-EMR-Version sbezeichnung	Hue-Version	Mit Hue installierte Komponent en
emr-6.15.0	Hue 4.11.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server,



Amazon-EMR-Version sbezeichnung	Hue-Version	Mit Hue installierte Komponenten
		mariadb-server, oozie-client, oozie-server

### Hue-Version für 5.36.2

Die folgende Tabelle listet die Version von Hue auf, die in der neuesten Version der Amazon-EMR-5.x-Serie enthalten ist, zusammen mit den Komponenten, die Amazon EMR mit Hue installiert.

Die Version der Komponenten, die in dieser Version mit Hue installiert wurden, finden Sie unter [Komponentenversionen von Version 5.36.2](#).

### Hue-Versionsinformationen für emr-5.36.2

Amazon-EMR-Version sbezeichnung	Hue-Version	Mit Hue installierte Komponenten
emr-5.36.2	Hue 4.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

## Unterstützte und nicht unterstützte Features von Hue in Amazon EMR

Mit Amazon EMR 7.0.0 und höher benötigt Hue Python 3.9 oder höher. Wenn Sie die Verschlüsselung während der Übertragung verwenden, benötigen Sie einen 2048-Bit-Verschlüsselungsschlüssel.

- Amazon S3 und Hadoop File System (HDFS) Browser
  - Wenn Sie über die entsprechenden Berechtigungen verfügen, können Sie Daten durchsuchen und zwischen dem flüchtigen HDFS-Speicher und S3-Buckets in Ihrem Konto verschieben.
  - Hauptbenutzer in Hue können standardmäßig alle auf Dateien zugreifen, auf die Amazon-EMR-IAM-Rollen zugreifen dürfen. Neu erstellte Benutzer verfügen nicht automatisch über Zugriffsberechtigungen auf den Amazon-S3-Filebrowser und für ihre Gruppen müssen die `filebrowser.s3_access`-Berechtigungen aktiviert sein.
- Hive – Führen Sie interaktive Abfragen für Ihre Daten aus. Hive kann außerdem für die programmgesteuerte Prototyp-Erstellung und für als Stapel verarbeitete Abfragen genutzt werden.
- Pig – Führen Sie Skripts für Ihre Daten aus oder nutzen Sie interaktive Befehle.
- Oozie – Erstellen und überwachen Sie Oozie-Workflows.
- Metastore-Manager – Mit dem Metastore-Manager können Sie die Inhalte des Hive-Metastores anzeigen und bearbeiten (importieren/erstellen, löschen usw.).
- Auftragsbrowser – Sehen Sie sich den Status Ihrer gesendeten Hadoop-Aufträge an.
- Benutzermanagement - Verwalten Sie Hue-Konten und integrieren Sie LDAP-Benutzer in Hue.
- AWS Beispiele — Es gibt mehrere "ready-to-run" Beispiele für die Verarbeitung von Beispieldaten aus verschiedenen AWS Diensten mithilfe von Anwendungen in Hue. Wenn Sie sich bei Hue anmelden, gelangen Sie zur Startseite der Hue-Anwendung, wo sich die vorinstallierten Beispiele befinden.
- Der Livy-Server wird nur in Amazon-EMR-Version 5.9.0 und höher unterstützt.
- Um Hue-Notebook für Spark zu verwenden, müssen Sie Hue mit Livy und Spark installieren.
- Trino/Presto – Hue unterstützt die Verbindung zu Trino- und Presto-Datenservern. Mit Amazon EMR 7.0.0 und höher wird diese Konnektivität mit `trino-python-client`- und `presto-python-client`-Konnektoren erleichtert. Versionen unter 7.0.0 verwenden einen JDBC-Konnektor über ein JDBC-Gateway.
- Das Hue-Dashboard wird nicht unterstützt.
- PostgreSQL wird nicht unterstützt.

## Überlegungen

Beachten Sie die folgenden Einschränkungen, wenn Sie [Hue](#) auf Amazon EMR verwenden.

## Leistung bei großen Hue-Metadatentabellen

Wenn die Hue-Metadaten-Datenbank zu groß wird, kann sich die Leistung verschlechtern. Um die Größe der Tabellen zu überprüfen, stellen Sie zunächst mit SSH eine Verbindung zum primären Knoten des Amazon EMR on Amazon EC2 EC2-Clusters her und führen Sie den Befehl `sudo mysql -u root` zum Starten der MySQL-CLI aus. Führen Sie die Abfrage aus, um die Größe Ihrer Tabelle zu ermitteln. `SELECT COUNT(*) FROM hue.<table_name>` Was das `<table_name>` sein kann, finden Sie im Folgenden:

- `desktop_dokument`
- `Desktop_Dokument2`
- `oozie_job`
- `beeswax_savedquery`
- `bienenwachs_session`
- Verlauf von `beeswax_query`

Wenn die Ausführung dieser Abfrage eine Anzahl von mehr als 100000 zurückgibt, sollten Sie den folgenden Bereinigungsbefehl ausführen, um die alten Datensätze zu löschen.

```
cd /opt/cloudera/parcels/CDH/lib/hue # Hue home directory ./build/env/bin/hue
desktop_document_cleanup
```

Weitere Informationen zur Bereinigung Ihrer Datenbank finden Sie auf der [Referenzseite in der Hue-Dokumentation](#).

Hue bereinigt die Tabellen nicht automatisch, aber Amazon EMR Versionen 5.12.0 und höher bieten eine Methode zum Löschen alter Dokumente in den Tabellen. Erstellen Sie das folgende Shell-Skript und führen Sie es als Schritt in einem Amazon EMR-Cluster mit einem Integer-Parameter aus, der die maximale Anzahl von Tagen angibt, für die Dokumente in der Metadaten-Datenbank aufbewahrt werden sollen.

```
#!/bin/bash
if grep isMaster /mnt/var/lib/info/instance.json | grep false;
then
    echo "This is not the primary node; do nothing, exiting"
    exit 0
fi
```

```
while [ ! -f /usr/lib/hue/desktop/core/src/desktop/management/commands/  
desktop_document_cleanup.py ]  
do  
    sleep 1  
done  
sudo systemctl stop hue.service  
sudo sed -i 's+ LOG.warn+ # LOG.warn+g' /usr/lib/hue/desktop/core/src/desktop/  
management/commands/desktop_document_cleanup.py  
sudo /usr/lib/hue/build/env/bin/hue desktop_document_cleanup --keep-days $1  
sudo systemctl start hue.service
```

## Inkompatibilität zwischen Hue-Versionen

Wenn du dieselbe Hue-Metadaten-Datenbank für mehrere Hue-fähige Cluster verwendest, empfehlen wir, dass auf diesen Clustern dieselbe Version von Hue ausgeführt wird. Verschiedene Versionen von Hue können unterschiedliche Schemas für die Hue-Metadaten-Datenbank haben. Die Verwendung derselben Datenbank für verschiedene Versionen kann dazu führen, dass eine Hue-Installation fehlschlägt. Wenn Sie beispielsweise versuchen, dieselbe Datenbank für zwei Cluster zu verwenden, auf denen 4.10.0 und 4.11.0 installiert sind, kann dies zu Anmeldefehlern für Benutzer führen, die versuchen, sich beim 4.10.0-Cluster anzumelden.

Eine Liste der Amazon EMR-Release-Labels und der entsprechenden installierten Versionen von [Hue finden Sie in der Hue-Versionshistorie](#).

## Verbindung zur Hue-Webbenutzeroberfläche herstellen

Das Herstellen von Hue entspricht der Verbindung mit einer beliebigen HTTP-Schnittstelle, die auf dem Hauptknoten eines Clusters gehostet wird. Im folgenden Verfahren wird beschrieben, wie Sie auf die Hue-Benutzeroberfläche zugreifen. Weitere Informationen finden Sie unter [Anzeigen von auf Amazon-EMR-Clustern gehosteten Webschnittstellen](#) im Verwaltungshandbuch für Amazon EMR.

So zeigen Sie die Hue-Webbenutzeroberfläche an

1. Befolgen Sie die folgenden Anweisungen zum [Einrichten eines SSH-Tunnels zum Hauptknoten mithilfe der dynamischen Port-Weiterleitung](#) im Verwaltungshandbuch für Amazon EMR.
2. Geben Sie die folgende Adresse in Ihren Browser ein, um die Hue-Webschnittstelle zu öffnen: `http://master public DNS:8888`, wobei *master public dns* der öffentliche DNS-Name Ihres Cluster-Master-Knotens ist, beispielsweise `ec2-11-22-333-44.compute-1.amazonaws.com`.

3. Wenn Sie sich zum ersten Mal als Administrator anmelden, geben Sie im Hue-Anmeldebildschirm einen Benutzernamen und ein Passwort ein, um Ihr Hue-Super-Konto einzurichten, und wählen Sie dann Konto erstellen aus. Geben Sie andernfalls Ihren Benutzernamen und Ihr Passwort ein, und wählen Sie Create account (Konto erstellen) aus bzw. geben Sie die Anmeldeinformationen ein, die Ihnen von Ihrem Administrator bereitgestellt wurden.

## Verwenden von Hue mit einer Remote-Datenbank in Amazon RDS

Standardmäßig werden Benutzerinformationen und Abfrageverläufe von Hue in einer lokalen MySQL-Datenbank auf dem Master-Knoten gespeichert. Sie können alternativ einen oder mehrere Hue-fähige Cluster mithilfe einer in Amazon S3 gespeicherten Konfiguration und einer MySQL-Datenbank in Amazon Relational Database Service (Amazon RDS) erstellen. So können Sie Benutzerdaten und Abfrageverläufe, die von Hue erstellt wurden, dauerhaft speichern, ohne dass der Amazon EMR-Cluster weiter ausgeführt wird. Wir empfehlen die Verwendung der serverseitigen Amazon-S3-Verschlüsselung zum Speichern der Konfigurationsdatei.

Erstellen Sie zunächst die Remote-Datenbank für Hue.


So erstellen Sie die externe MySQL-Datenbank

1. Öffnen Sie die Amazon-RDS-Konsole unter <https://console.aws.amazon.com/rds/>.
2. Klicken Sie auf Launch a DB Instance (Eine DB-Instance starten).
3. Wählen Sie „MySQL“ aus, und klicken Sie auf Select (Auswählen).
4. Übernehmen Sie die Standardauswahl von Multi-AZ Deployment and Provisioned IOPS Storage (Multi-AZ-Bereitstellung und bereitgestellte IOPS-Speicherung), und klicken Sie auf Next (Weiter).
5. Übernehmen Sie die Standardwerte für die Instance-Spezifikationen, geben Sie Einstellungen an, und klicken Sie auf Next (Weiter).
6. Wählen Sie auf der Seite "Configure Advanced Settings" eine geeignete Sicherheitsgruppe und einen Datenbanknamen aus. Die Sicherheitsgruppe, die Sie verwenden, muss mindestens eingehenden TCP-Zugriff für Port 3306 über den Master-Knoten Ihres Clusters zulassen. Wenn Sie zu diesem Zeitpunkt noch keinen Cluster erstellt haben, können Sie für alle Hosts eine Verbindung mit Port 3306 zulassen und die Sicherheitsgruppe anpassen, nachdem Sie den Cluster gestartet haben. Klicken Sie auf Launch DB Instance (DB-Instance starten).

- Wählen Sie im RDS-Dashboard Instances und die Instance aus, die Sie soeben erstellt haben. Wenn Ihre Datenbank verfügbar ist, notieren Sie dbname, Benutzername, Kennwort und RDS-Instance. Diese Informationen benötigen Sie zum Erstellen und Konfigurieren Ihres Clusters.


Um eine externe MySQL-Datenbank für Hue anzugeben, wenn Sie einen Cluster mit dem starten AWS CLI

Um beim Starten eines Clusters mit dem eine externe MySQL-Datenbank für Hue anzugeben AWS CLI, verwenden Sie die Informationen, die Sie bei der Erstellung Ihrer RDS-Instanz für die Konfiguration `hue.ini` mit einem Konfigurationsobjekt notiert haben.

 Note

Sie können mehrere Cluster mit derselben externen Datenbank erstellen, aber jeder Cluster gibt Abfrageverläufe und Benutzerdaten frei.

- Erstellen Sie mit dem einen Cluster AWS CLI, auf dem Hue installiert ist, verwenden Sie die externe Datenbank, die Sie erstellt haben, und verweisen Sie auf eine Konfigurationsdatei mit einer Konfigurationsklassifizierung für Hue, die die Datenbankeigenschaften spezifiziert. Im folgenden Beispiel wird ein Cluster mit installiertem Hue erstellt, der auf die Konfigurationsdatei Amazon S3 in `myConfig.json` verweist, die die Datenbankkonfiguration angibt.

 Note

Linux-Zeilenfortsetzungszeichen (`\`) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (`^`).

```
aws emr create-cluster --release-label emr-7.1.0 --applications Name=Hue Name=Spark
Name=Hive \
--instance-type m5.xlarge --instance-count 3 \
--configurations https://s3.amazonaws.com/mybucket/myfolder/myConfig.json --use-
default-roles
```

Beispielinhalte der `myConfig.json`-Datei werden unten angezeigt. Ersetzen Sie *dbname*, *username*, *password* und *RDS instance hostname* durch die Werte, die Sie zuvor im RDS-Dashboard angegeben haben.

```
[{
  "Classification": "hue-ini",
  "Properties": {},
  "Configurations": [
    {
      "Classification": "desktop",
      "Properties": {},
      "Configurations": [
        {
          "Classification": "database",
          "Properties": {
            "name": "dbname",
            "user": "username",
            "password": "password",
            "host": "RDS instance hostname",
            "port": "3306",
            "engine": "mysql"
          },
          "Configurations": []
        }
      ]
    }
  ]
}]
```

## Fehlerbehebung

Im Falle eines Amazon RDS-Failovers

Beim Ausführen einer Abfrage können Verzögerungen auftreten, da die Hue-Datenbank-Instance nicht reagiert oder sich in einem Failover befindet. Im Folgenden finden Sie einige Fakten und Leitlinien für dieses Problem:

- Wenn Sie sich bei der Amazon-RDS-Konsole anmelden, können Sie nach Failover-Ereignissen suchen. Um beispielsweise festzustellen, ob ein Failover ausgeführt wird oder aufgetreten ist,

suchen Sie nach Ereignissen wie "Multi-AZ instance failover started" und "Multi-AZ instance failover completed".

- Es dauert ca. 30 Sekunden, bis eine RDS-Instance einen Failover ausgeführt hat.
- Wenn Sie longer-than-normal Antworten auf Anfragen in Hue erhalten, versuchen Sie, die Abfrage erneut auszuführen.

## Erweiterte Konfigurationen für Hue

In diesem Abschnitt werden folgende Themen behandelt.

Themen

- [Konfigurieren von Hue für LDAP-Benutzer](#)

### Konfigurieren von Hue für LDAP-Benutzer

Dank der Integration mit LDAP können Benutzer sich bei Hue mit vorhandenen Anmeldeinformationen anmelden, die in einem LDAP-Verzeichnis gespeichert sind. Wenn Sie Hue mit LDAP integrieren, müssen Sie Benutzerinformationen nicht separat in Hue verwalten. Die Informationen unten zeigen die Hue-Integration mit Microsoft Active Directory, die Konfigurationsoptionen sind jedoch mit jedem beliebigen LDAP-Verzeichnis vergleichbar.

Die LDAP-Authentifizierung wird zunächst an den Server gebunden und stellt die Verbindung her. Die hergestellte Verbindung wird für alle nachfolgenden Abfragen zum Suchen nach LDAP-Benutzerinformationen verwendet. Wenn Ihr Active Directory-Server keine anonymen Verbindungen zulässt, muss eine Verbindung mit einem definierten Bind-Namen und -Passwort hergestellt werden. Der definierte Bind-Name (DN) wird durch die Konfigurationseinstellung `bind_dn` definiert. Das Bind-Passwort wird von der Konfigurationseinstellung `bind_password` definiert. Hue hat zwei Möglichkeiten für die Verknüpfung von LDAP-Anfragen: Search Bind und Direct Bind. Die bevorzugte Methode für die Verwendung von Hue mit Amazon EMR ist Search Bind.

Wenn Search Bind mit Active Directory verwendet wird, zieht Hue das Benutzernamenattribut (definiert durch `user_name_attr config`) heran, um das Attribut zu finden, das vom definierten Base-Namen (DN) abgerufen werden muss. Search Bind ist hilfreich, wenn der vollständige DN für den Hue-Benutzer nicht bekannt ist.

Es ist beispielsweise möglich, dass Sie `user_name_attr config` auf den allgemeinen Namen (oder CN) festgelegt haben. In diesem Fall verwendet der Active Directory-Server den Hue-



Benutzernamen, der während der Anmeldung angegeben wurde, um die Verzeichnisstruktur nach einem übereinstimmenden allgemeinen Namen zu durchsuchen, und zwar beginnend mit dem definierten Base-Namen. Wenn er den allgemeinen Namen für den Hue-Benutzer findet, gibt er den definierten Namen zurück. Hue erstellt dann einen definierten Namen, der für die Authentifizierung des Benutzers durch eine Bind-Operation verwendet wird.

#### Note

Search Bind durchsucht die Benutzernamen in allen Verzeichnis-Unterstrukturen beginnend mit dem definierten Namen. Der in der Hue-LDAP-Konfiguration angegebene definierte Base-Name sollte das nächste übergeordnete Element des Benutzernamens sein, andernfalls wirkt sich dies negativ auf die Leistung der LDAP-Authentifizierung aus.

Wenn Direct Bind mit Active Directory verwendet wird, muss die exakte `nt_domain` oder `ldap_username_pattern` für die Authentifizierung verwendet werden. Wenn bei der Verwendung von Direct Bind das NT-Domainattribut (definiert von der Konfigurationseinstellung `nt_domain`) definiert wird, wird eine Vorlage für den definierten Benutzernamen im folgenden Format erstellt: `<login username>@nt_domain`. Diese Vorlage wird verwendet, um alle Verzeichnis-Unterstrukturen beginnend mit dem definierten Namen zu durchsuchen. Wenn die NT-Domain nicht konfiguriert ist, sucht Hue nach einem exakten DN-Muster für den Benutzer (definiert durch die Konfigurationseinstellung `ldap_username_pattern`). In diesem Fall sucht der Server nach einem passenden `ldap_username_pattern`-Wert in allen Verzeichnis-Unterstrukturen beginnend mit dem definierten Base-Namen.

So starten Sie einen Cluster mit LDAP-Eigenschaften für Hue mithilfe der AWS CLI

- Zum Festlegen von LDAP-Eigenschaften für `hue.ini` erstellen Sie einen Cluster mit installiertem Hue und verweisen auf eine JSON-Datei mit Konfigurationseigenschaften für LDAP. Unten wird ein Beispielbefehl angezeigt, der auf eine Konfigurationsdatei `myConfig.json` verweist, die Amazon S3 gespeichert ist.

```
aws emr create-cluster --release-label emr-7.1.0 --applications Name=Hue Name=Spark
  Name=Hive \
--instance-type m5.xlarge --instance-count 3 --configurations https://
s3.amazonaws.com/mybucket/myfolder/myConfig.json.
```

Beispielinhalte von `myConfig.json` werden unten angezeigt.

```
[
  {
    "Classification": "hue-ini",
    "Properties": {},
    "Configurations": [
      {
        "Classification": "desktop",
        "Properties": {},
        "Configurations": [
          {
            "Classification": "ldap",
            "Properties": {},
            "Configurations": [
              {
                "Classification": "ldap_servers",
                "Properties": {},
                "Configurations": [
                  {
                    "Classification": "yourcompany",
                    "Properties": {
                      "base_dn":
"DC=yourcompany,DC=hue,DC=com",
                      "ldap_url": "ldap://ldapurl",
                      "search_bind_authentication": "true",
                      "bind_dn":
"CN=hue,CN=users,DC=yourcompany,DC=hue,DC=com",
                      "bind_password": "password"
                    },
                    "Configurations": []
                  }
                ]
              }
            ]
          }
        ],
        "Properties": {
          "backend": "desktop.auth.backend.LdapBackend"
        }
      }
    ]
  }
]
```

```
]
  }
]
```

### Note

Ab Amazon-EMR-Version 5.21.0 können Sie Cluster-Konfigurationen überschreiben und zusätzliche Konfigurationsklassifikationen für jede Instance-Gruppe in einem ausgeführten Cluster angeben. Dazu verwenden Sie die Amazon EMR-Konsole, das AWS Command Line Interface (AWS CLI) oder das AWS SDK. Weitere Informationen finden Sie unter [Angeben einer Konfiguration für eine Instance-Gruppe in einem aktiven Cluster](#).

So zeigen Sie LDAP-Einstellungen in Hue an

1. Überprüfen Sie, ob Sie über eine aktive VPN-Verbindung oder einen aktiven SSH-Tunnel zum Hauptknoten des Amazon-EMR-Clusters verfügen. Geben Sie dann in Ihrem Browser :8888 ein, um die *master-public-dns*Hue-Weboberfläche zu öffnen.
2. Melden Sie sich mit Ihren Hue-Administratoranmeldeinformationen an. Klicken Sie im Dialogfeld Did you know? (Wussten Sie schon?) auf Got it, prof! (Verstanden!) um es zu schließen.
3. Klicken Sie auf das Symbol Hue in der Symbolleiste.
4. Wählen Sie auf der Seite About Hue (Über Hue) auf Configuration (Konfiguration) aus.
5. Klicken Sie im Abschnitt Configuration Sections and Variables (Konfigurationsabschnitte und -variablen) auf Desktop.
6. Scrollen Sie zum Abschnitt LDAP, um Ihre Einstellungen anzuzeigen.

## Hue-Versionsverlauf

In der folgenden Tabelle sind die Version von Hue aufgeführt, die in jeder Release-Version von Amazon EMR enthalten ist, zusammen mit den Komponenten, die mit der Anwendung installiert wurden. Informationen zu den Komponentenversionen in den einzelnen Versionen finden Sie im Abschnitt Komponentenversion für Ihre Version in [Amazon-EMR-7.x-Versionen](#), [Amazon-EMR-6.x-Versionen](#) oder [Amazon-EMR-5.x-Versionen](#).

## Die Hue Versionsinformationen

Amazon-EMR-Version sbezeichnung	Hue-Version	Mit Hue installierte Komponent en
emr-5.36.2	4.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-7.1.0	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-7.0.0	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server,

Amazon-EMR-Version sbezeichnung	Hue-Version	Mit Hue installierte Komponent en
		mariadb-server, oozie-client, oozie-server
emr-6.15.0	4.11.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-6.14.0	4.11.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Amazon-EMR-Version sbezeichnung	Hue-Version	Mit Hue installierte Komponent en
emr-6.13.0	4.11.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-6.12.0	4.11.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Amazon-EMR-Version sbezeichnung	Hue-Version	Mit Hue installierte Komponent en
emr-6.11.1	4.11.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-6.11.0	4.11.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Amazon-EMR-Version sbezeichnung	Hue-Version	Mit Hue installierte Komponent en
emr-6.10.1	4.10.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-6.10.0	4.10.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mariadb-server, oozie-client, oozie-server



Amazon-EMR-Version sbezeichnung	Hue-Version	Mit Hue installierte Komponent en
emr-6.9.1	4.10.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-6.9.0	4.10.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Amazon-EMR-Version sbezeichnung	Hue-Version	Mit Hue installierte Komponent en
emr-6.8.1	4.10.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-6.8.0	4.10.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Amazon-EMR-Version sbezeichnung	Hue-Version	Mit Hue installierte Komponent en
emr-6.7.0	4.10.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-5.36.1	4.10.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Amazon-EMR-Version sbezeichnung	Hue-Version	Mit Hue installierte Komponent en
emr-5.36.0	4.10.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-6.6.0	4.10.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Amazon-EMR-Version sbezeichnung	Hue-Version	Mit Hue installierte Komponent en
emr-5.35.0	4.10.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-6.5.0	4.9.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Amazon-EMR-Version sbezeichnung	Hue-Version	Mit Hue installierte Komponent en
emr-6.4.0	4.9.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-6.3.1	4.9.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Amazon-EMR-Version sbezeichnung	Hue-Version	Mit Hue installierte Komponent en
emr-6.3.0	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-6.2.1	4.8.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Amazon-EMR-Version sbezeichnung	Hue-Version	Mit Hue installierte Komponent en
emr-6.2.0	4.8.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-6.1.1	4.7.1	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mariadb-server, oozie-client, oozie-server



Amazon-EMR-Version sbezeichnung	Hue-Version	Mit Hue installierte Komponent en
emr-6.1.0	4.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-6.0.1	4.4.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Amazon-EMR-Version sbezeichnung	Hue-Version	Mit Hue installierte Komponent en
emr-6.0.0	4.4.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-5.34.0	4.9.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Amazon-EMR-Version sbezeichnung	Hue-Version	Mit Hue installierte Komponent en
emr-5.33.1	4.9.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-5.33.0	4.9.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Amazon-EMR-Version sbezeichnung	Hue-Version	Mit Hue installierte Komponent en
emr-5.32.1	4.8.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-5.32.0	4.8.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Amazon-EMR-Version sbezeichnung	Hue-Version	Mit Hue installierte Komponent en
emr-5.31.1	4.7.1	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-5.31.0	4.7.1	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Amazon-EMR-Version sbezeichnung	Hue-Version	Mit Hue installierte Komponent en
emr-5.30.2	4.6.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-5.30.1	4.6.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Amazon-EMR-Version sbezeichnung	Hue-Version	Mit Hue installierte Komponent en
emr-5.30.0	4.6.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-5.29.0	4.4.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Amazon-EMR-Version sbezeichnung	Hue-Version	Mit Hue installierte Komponent en
emr-5.28.1	4.4.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.28.0	4.4.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mysql-server, oozie-client, oozie-server



Amazon-EMR-Version sbezeichnung	Hue-Version	Mit Hue installierte Komponent en
emr-5.27.1	4.4.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.27.0	4.4.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Amazon-EMR-Version sbezeichnung	Hue-Version	Mit Hue installierte Komponent en
emr-5.26.0	4.4.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.25.0	4.4.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Amazon-EMR-Version sbezeichnung	Hue-Version	Mit Hue installierte Komponent en
emr-5.24.1	4.4.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.24.0	4.4.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Amazon-EMR-Version sbezeichnung	Hue-Version	Mit Hue installierte Komponent en
emr-5.23.1	4.3.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.23.0	4.3.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Amazon-EMR-Version sbezeichnung	Hue-Version	Mit Hue installierte Komponent en
emr-5.22.0	4.3.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.21.2	4.3.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Amazon-EMR-Version sbezeichnung	Hue-Version	Mit Hue installierte Komponent en
emr-5.21.1	4.3.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.21.0	4.3.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Amazon-EMR-Version sbezeichnung	Hue-Version	Mit Hue installierte Komponent en
emr-5.20.1	4.3.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.20.0	4.3.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Amazon-EMR-Version sbezeichnung	Hue-Version	Mit Hue installierte Komponent en
emr-5.19.1	4.2.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.19.0	4.2.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mysql-server, oozie-client, oozie-server



Amazon-EMR-Version sbezeichnung	Hue-Version	Mit Hue installierte Komponent en
emr-5.18.1	4.2.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.18.0	4.2.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Amazon-EMR-Version sbezeichnung	Hue-Version	Mit Hue installierte Komponent en
emr-5.17.2	4.2.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.17.1	4.2.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Amazon-EMR-Version sbezeichnung	Hue-Version	Mit Hue installierte Komponent en
emr-5.17.0	4.2.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.16.1	4.2.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Amazon-EMR-Version sbezeichnung	Hue-Version	Mit Hue installierte Komponent en
emr-5.16.0	4.2.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.15.1	4.2.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Amazon-EMR-Version sbezeichnung	Hue-Version	Mit Hue installierte Komponent en
emr-5.15.0	4.2.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.14.2	4.1.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Amazon-EMR-Version sbezeichnung	Hue-Version	Mit Hue installierte Komponent en
emr-5.14.1	4.1.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.14.0	4.1.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Amazon-EMR-Version sbezeichnung	Hue-Version	Mit Hue installierte Komponent en
emr-5.13.1	4.1.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.13.0	4.1.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Amazon-EMR-Version sbezeichnung	Hue-Version	Mit Hue installierte Komponent en
emr-5.12.3	4.1.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.12.2	4.1.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mysql-server, oozie-client, oozie-server



Amazon-EMR-Version sbezeichnung	Hue-Version	Mit Hue installierte Komponent en
emr-5.12.1	4.1.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.12.0	4.1.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Amazon-EMR-Version sbezeichnung	Hue-Version	Mit Hue installierte Komponent en
emr-5.11.4	4.0.1	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.11.3	4.0.1	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Amazon-EMR-Version sbezeichnung	Hue-Version	Mit Hue installierte Komponent en
emr-5.11.2	4.0.1	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.11.1	4.0.1	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Amazon-EMR-Version sbezeichnung	Hue-Version	Mit Hue installierte Komponent en
emr-5.11.0	4.0.1	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.10.1	4.0.1	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Amazon-EMR-Version sbezeichnung	Hue-Version	Mit Hue installierte Komponent en
emr-5.10.0	4.0.1	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.9.1	4.0.1	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Amazon-EMR-Version sbezeichnung	Hue-Version	Mit Hue installierte Komponent en
emr-5.9.0	4.0.1	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.8.3	3.12.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Amazon-EMR-Version sbezeichnung	Hue-Version	Mit Hue installierte Komponent en
emr-5.8.2	3.12.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.8.1	3.12.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Amazon-EMR-Version sbezeichnung	Hue-Version	Mit Hue installierte Komponent en
emr-5.8.0	3.12.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.7.1	3.12.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mysql-server, oozie-client, oozie-server



Amazon-EMR-Version sbezeichnung	Hue-Version	Mit Hue installierte Komponent en
emr-5.7.0	3.12.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.6.1	3.12.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Amazon-EMR-Version sbezeichnung	Hue-Version	Mit Hue installierte Komponent en
emr-5.6.0	3.12.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.5.4	3.12.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hue-serve r, mysql-server, oozie-client, oozie-server
emr-5.5.3	3.12.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hue-serve r, mysql-server, oozie-client, oozie-server

Amazon-EMR-Version sbezeichnung	Hue-Version	Mit Hue installierte Komponent en
emr-5.5.2	3.12.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hue-serve r, mysql-server, oozie-client, oozie-server
emr-5.5.1	3.12.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hue-serve r, mysql-server, oozie-client, oozie-server
emr-5.5.0	3.12.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hue-serve r, mysql-server, oozie-client, oozie-server

Amazon-EMR-Version sbezeichnung	Hue-Version	Mit Hue installierte Komponent en
emr-5.4.1	3.11.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hue-serve r, mysql-server, oozie-client, oozie-server
emr-5.4.0	3.11.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hue-serve r, mysql-server, oozie-client, oozie-server
emr-5.3.2	3.11.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hue-serve r, mysql-server, oozie-client, oozie-server

Amazon-EMR-Version sbezeichnung	Hue-Version	Mit Hue installierte Komponent en
emr-5.3.1	3.11.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hue-serve r, mysql-server, oozie-client, oozie-server
emr-5.3.0	3.11.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hue-serve r, mysql-server, oozie-client, oozie-server
emr-5.2.3	3.10.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hue-serve r, mysql-server, oozie-client, oozie-server

Amazon-EMR-Version sbezeichnung	Hue-Version	Mit Hue installierte Komponent en
emr-5.2.2	3.10.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hue-serve r, mysql-server, oozie-client, oozie-server
emr-5.2.1	3.10.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hue-serve r, mysql-server, oozie-client, oozie-server
emr-5.2.0	3.10.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hue-serve r, mysql-server, oozie-client, oozie-server

Amazon-EMR-Version sbezeichnung	Hue-Version	Mit Hue installierte Komponent en
emr-5.1.1	3.10.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hue-serve r, mysql-server, oozie-client, oozie-server
emr-5.1.0	3.10.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hue-serve r, mysql-server, oozie-client, oozie-server
emr-5.0.3	3.10.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hue-server, mysql-server, oozie-server

Amazon-EMR-Version sbezeichnung	Hue-Version	Mit Hue installierte Komponent en
emr-5.0.2	3.10.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hue-server, mysql-server, oozie-server
emr-5.0.1	3.10.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hue-server, mysql-server, oozie-server
emr-5.0.0	3.10.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hue-server, mysql-server, oozie-server



Amazon-EMR-Version sbezeichnung	Hue-Version	Mit Hue installierte Komponent en
emr-4.9.6	3.7.1	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hue-serve r, mysql-server, oozie-client, oozie-server
emr-4.9.5	3.7.1	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hue-serve r, mysql-server, oozie-client, oozie-server
emr-4.9.4	3.7.1	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hue-serve r, mysql-server, oozie-client, oozie-server

Amazon-EMR-Version sbezeichnung	Hue-Version	Mit Hue installierte Komponent en
emr-4.9.3	3.7.1	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hue-serve r, mysql-server, oozie-client, oozie-server
emr-4.9.2	3.7.1	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hue-serve r, mysql-server, oozie-client, oozie-server
emr-4.9.1	3.7.1	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hue-serve r, mysql-server, oozie-client, oozie-server

Amazon-EMR-Version sbezeichnung	Hue-Version	Mit Hue installierte Komponent en
emr-4.8.5	3.7.1	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hue-serve r, mysql-server, oozie-client, oozie-server
emr-4.8.4	3.7.1	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hue-serve r, mysql-server, oozie-client, oozie-server
emr-4.8.3	3.7.1	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hue-serve r, mysql-server, oozie-client, oozie-server

Amazon-EMR-Version sbezeichnung	Hue-Version	Mit Hue installierte Komponent en
emr-4.8.2	3.7.1	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hue-serve r, mysql-server, oozie-client, oozie-server
emr-4.8.1	3.7.1	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hue-serve r, mysql-server, oozie-client, oozie-server
emr-4.8.0	3.7.1	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hue-serve r, mysql-server, oozie-client, oozie-server

Amazon-EMR-Version sbezeichnung	Hue-Version	Mit Hue installierte Komponent en
emr-4.7.4	3.7.1	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hue-serve r, mysql-server, oozie-client, oozie-server
emr-4.7.3	3.7.1	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hue-serve r, mysql-server, oozie-client, oozie-server
emr-4.7.2	3.7.1	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hue-serve r, mysql-server, oozie-client, oozie-server

Amazon-EMR-Version sbezeichnung	Hue-Version	Mit Hue installierte Komponent en
emr-4.7.1	3.7.1	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hue-server, mysql-server, oozie-server
emr-4.7.0	3.7.1	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hue-server, mysql-server, oozie-server
emr-4.6.1	3.7.1	emrfs, emr-ddb, emr-goodi es, emr-kinesis, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hue-server, mysql-server, oozie-server

Amazon-EMR-Version sbezeichnung	Hue-Version	Mit Hue installierte Komponent en
emr-4.6.0	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-server
emr-4.5.0	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-server
emr-4.4.0	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-server

Amazon-EMR-Version sbezeichnung	Hue-Version	Mit Hue installierte Komponent en
emr-4.3.0	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-server
emr-4.2.0	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-server
emr-4.1.0	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-server



# Iceberg

[Apache Iceberg](#) ist ein offenes Tabellenformat für große Datensätze in Amazon Simple Storage Service (Amazon S3). Es bietet schnelle Abfrageleistung bei großen Tabellen, atomare Commits, gleichzeitige Schreibvorgänge und eine SQL-kompatible Tabellenentwicklung. Ab Amazon EMR 6.5.0 können Sie Apache Spark 3 auf Amazon-EMR-Clustern mit dem Iceberg-Tabellenformat verwenden.

Die folgende Tabelle listet die Version von Iceberg auf, die in der neuesten Version der Amazon-EMR-7.x-Serie enthalten ist, zusammen mit den Komponenten, die Amazon EMR mit Iceberg installiert.

Informationen zur Version der Komponenten, die in dieser Version mit Iceberg installiert wurden, finden Sie unter [Komponentenversionen von Version 7.1.0](#).

Iceberg-Versionsinformationen für emr-7.1.0

Amazon-EMR-Version sbezeichnung	Iceberg-Version	Mit Iceberg installierte Komponenten
emr-7.1.0	Iceberg 1.4.3-amzn-0	Not available.

Die folgende Tabelle listet die Version von Iceberg auf, die in der neuesten Version der Amazon-EMR-6.x-Serie enthalten ist, zusammen mit den Komponenten, die Amazon EMR mit Iceberg installiert.

Die Version der Komponenten, die mit Iceberg in dieser Version installiert wurden, finden Sie unter [Komponentenversionen der Version 6.15.0](#).

Iceberg-Versionsinformationen für emr-6.15.0

Amazon-EMR-Version sbezeichnung	Iceberg-Version	Mit Iceberg installierte Komponenten
emr-6.15.0	Iceberg 1.4.0-amzn-0	Not available.

Themen

- [Funktionsweise von Iceberg](#)

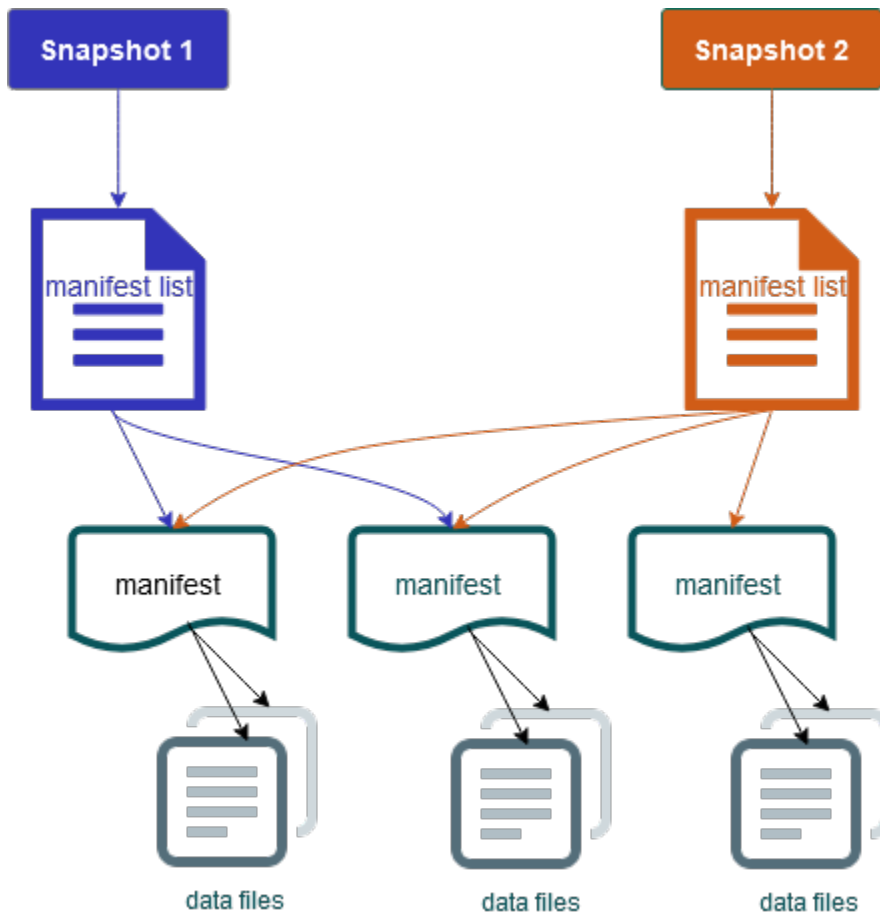
- [Einen Cluster verwenden, auf dem Iceberg installiert ist](#)
- [Überlegungen und Einschränkungen für die Verwendung von Iceberg in Amazon EMR](#)
- [Versionsverlauf von Iceberg](#)

## Funktionsweise von Iceberg

Iceberg verfolgt einzelne Datendateien in einer Tabelle statt in Verzeichnissen. Auf diese Weise können Autoren Datendateien an Ort und Stelle erstellen (Dateien werden nicht verschoben oder geändert). Außerdem können Autoren der Tabelle nur in einem expliziten Commit Dateien hinzufügen. Der Tabellenstatus wird in Metadatendateien beibehalten. Bei allen Änderungen am Tabellenstatus wird eine neue Metadatendatei erstellt, die die älteren Metadaten atomar ersetzt. Die Tabellenmetadatendatei verfolgt das Tabellenschema, die Partitionierungskonfiguration und andere Eigenschaften.

Sie enthält auch Snapshots des Tabelleninhalts. Jeder Snapshot ist ein vollständiger Satz von Datendateien in der Tabelle zu einem bestimmten Zeitpunkt. Snapshots sind in der Metadatendatei aufgeführt, aber die Dateien in einem Snapshot werden in separaten Manifestdateien gespeichert. Die atomaren Übergänge von einer Tabellenmetadatendatei zur nächsten ermöglichen die Isolierung von Snapshots. Leser verwenden den Snapshot, der aktuell war, als sie die Tabellenmetadaten geladen haben. Leser sind von Änderungen erst betroffen, wenn sie den Vorgang aktualisieren und einen neuen Speicherort für Metadaten auswählen. Datendateien in Snapshots werden in einer oder mehreren Manifestdateien gespeichert, die eine Zeile für jede Datendatei in der Tabelle, ihre Partitionsdaten und ihre Metriken enthalten. Ein Snapshot ist die Vereinigung aller Dateien in ihren Manifesten. Manifestdateien können auch von Snapshots gemeinsam genutzt werden, um zu vermeiden, dass Metadaten, die sich selten ändern, neu geschrieben werden.

Iceberg-Snapshot-Diagramm



Iceberg bietet folgende Features:

- Unterstützt ACID-Transaktionen und Zeitreisen in Ihrem Amazon-S3-Data-Lake.
- Commit-Wiederholungen profitieren von den Leistungsvorteilen [optimistischer Parallelität](#).
- Die Konfliktlösung auf Dateiebene führt zu einer hohen Parallelität.
- Mit Min-Max-Statistiken pro Spalte in Metadaten können Sie Dateien überspringen, was die Leistung bei selektiven Abfragen steigert.
- Sie können Tabellen in flexiblen Partitionslayouts organisieren, wobei die Partitionsentwicklung Aktualisierungen von Partitionsschemas ermöglicht. Abfragen und Datenvolumen können sich dann ändern, ohne auf physische Verzeichnisse angewiesen zu sein.
- Unterstützt die [Weiterentwicklung und Durchsetzung von Schemas](#).
- Iceberg-Tabellen dienen als idempotente Senken und widerspielbare Quellen. Dies ermöglicht Streaming und Batch-Support mit exakt einmal verwendeten Pipelines. Idempotente Senken verfolgen Schreiboperationen, die in der Vergangenheit erfolgreich waren. Daher kann die Senke

im Falle eines Fehlers erneut Daten anfordern und Daten löschen, wenn sie mehrfach gesendet wurden.

- Zeigen Sie Verlauf und Herkunft an, einschließlich Tabellenentwicklung, Operationsverlauf und Statistiken für jeden Commit.
- Migrieren Sie von einem vorhandenen Datensatz mit einer Auswahl an Datenformaten (Parquet, ORC, Avro) und Analyse-Engine (Spark, Trino, PrestoDB, Flink, Hive).

## Einen Cluster verwenden, auf dem Iceberg installiert ist

Dieser Abschnitt enthält Informationen zur Verwendung von Iceberg mit Spark, Trino, Flink und Hive.

### Verwenden Sie einen Iceberg-Cluster mit Spark

Ab Amazon-EMR-Version 6.5.0 können Sie Iceberg mit Ihrem Spark-Cluster verwenden, ohne dass Bootstrap-Aktionen erforderlich sind. Für Amazon-EMR-Versionen 6.4.0 und früher können Sie eine Bootstrap-Aktion verwenden, um alle erforderlichen Abhängigkeiten vorab zu installieren.

In diesem Tutorial verwenden Sie den, AWS CLI um mit Iceberg auf einem Amazon EMR Spark-Cluster zu arbeiten. Um die Konsole zur Erstellung eines Clusters mit Iceberg zu verwenden, folgen Sie den Schritten unter [Ein Data Lake von Apache Iceberg mit Amazon Athena, Amazon EMR und AWS Glue erstellen](#).

### Erstellen Sie einen Iceberg-Cluster

Sie können einen Cluster mit installiertem Iceberg mithilfe der AWS Management Console, der AWS CLI oder der Amazon EMR-API erstellen. In diesem Tutorial verwenden Sie den, AWS CLI um mit Iceberg auf einem Amazon EMR-Cluster zu arbeiten. Um die Konsole zur Erstellung eines Clusters mit Iceberg zu verwenden, folgen Sie den Schritten unter [Ein Data Lake von Apache Iceberg mit Amazon Athena, Amazon EMR und AWS Glue erstellen](#).

Um Iceberg auf Amazon EMR mit dem zu verwenden AWS CLI, erstellen Sie zunächst einen Cluster mit den folgenden Schritten. Informationen zur Spezifizierung der Iceberg-Klassifizierung mithilfe von finden Sie unter oder AWS CLI. [Geben Sie AWS CLI beim Erstellen eines Clusters eine Konfiguration an, indem Sie Beim Erstellen eines Clusters eine Konfiguration mit dem Java SDK angeben](#)

1. Erstellen Sie eine Datei `configurations.json`, mit folgendem Inhalt:

```
[{
  "Classification": "iceberg-defaults",
```

```
"Properties":{"iceberg.enabled":"true"}
}]
```

2. Im nächsten Schritt erstellen Sie einen Cluster mit der folgenden Konfiguration. Ersetzen Sie den Amazon-S3-Beispiel-Bucket-Pfad und die Subnetz-ID durch Ihre eigene.

```
aws emr create-cluster --release-label emr-6.5.0 \
--applications Name=Spark \
--configurations file://iceberg_configurations.json \
--region us-east-1 \
--name My_Spark_Iceberg_Cluster \
--log-uri s3://DOC-EXAMPLE-BUCKET/ \
--instance-type m5.xlarge \
--instance-count 2 \
--service-role EMR_DefaultRole_V2 \
--ec2-attributes
InstanceProfile=EMR_EC2_DefaultRole,SubnetId=subnet-1234567890abcdef0
```

Alternativ können Sie einen Amazon-EMR-Cluster mit der Spark-Anwendung erstellen und die Datei `/usr/share/aws/iceberg/lib/iceberg-spark3-runtime.jar` als JAR-Abhängigkeit in einen Spark-Auftrag aufnehmen. Weitere Informationen finden Sie unter [Anwendungen einreichen](#).

Um das JAR als Abhängigkeit in einen Spark-Auftrag aufzunehmen, fügen Sie der Spark-Anwendung die folgende Konfigurationseigenschaft hinzu:

```
--conf "spark.jars=/usr/share/aws/iceberg/lib/iceberg-spark3-runtime.jar"
```

Weitere Informationen zu den Abhängigkeiten von Spark-Aufträgen finden Sie unter [Abhängigkeitsverwaltung](#) im Apache-Spark-Dokument [Ausführen von Spark in Kubernetes](#).

## Initialisieren Sie eine Spark-Sitzung für Iceberg

In den folgenden Beispielen wird veranschaulicht, wie Sie die interaktive Spark-Shell starten und die Spark-Übermittlung und Amazon EMR Notebooks verwenden, um mit Iceberg in Amazon EMR zu arbeiten.

### spark-shell

1. Verbinden mit dem Master-Knoten über SSH. Weitere Informationen finden Sie unter [Mit SSH eine Verbindung zum Hauptknoten herstellen](#) im Verwaltungshandbuch für Amazon EMR.

2. Geben Sie den folgenden Befehl ein, um die Spark-Shell zu starten. Um die PySpark Shell zu verwenden, ersetzen Sie `spark-shell` durch `spark-shell`.

```
spark-shell \
--conf
  "spark.sql.extensions=org.apache.iceberg.spark.extensions.IcebergSparkSessionExtensions" \
--conf "spark.sql.catalog.dev=org.apache.iceberg.spark.SparkCatalog" \
--conf "spark.sql.catalog.dev.type=hadoop" \
--conf "spark.sql.catalog.dev.warehouse=s3://DOC-EXAMPLE-BUCKET/example-prefix/"
```

## spark-submit

1. Verbinden mit dem Master-Knoten über SSH. Weitere Informationen finden Sie unter [Mit SSH eine Verbindung zum Hauptknoten herstellen](#) im Verwaltungshandbuch für Amazon EMR.
2. Geben Sie den folgenden Befehl ein, um die Spark-Sitzung für Iceberg zu starten.

```
spark-submit \
--conf
  "spark.sql.extensions=org.apache.iceberg.spark.extensions.IcebergSparkSessionExtensions" \
--conf "spark.sql.catalog.dev=org.apache.iceberg.spark.SparkCatalog" \
--conf "spark.sql.catalog.dev.type=hadoop" \
--conf "spark.sql.catalog.dev.warehouse=s3://DOC-EXAMPLE-BUCKET/example-prefix/"
```

## EMR Studio notebooks

Um eine Spark-Sitzung mit EMR Studio-Notebooks zu initialisieren, konfigurieren Sie Ihre Spark-Sitzung mit dem `%%configure` magischen Befehl in Ihrem Amazon-EMR-Notebook, wie im folgenden Beispiel. Weitere Informationen finden Sie unter [Verwendung von EMR-Notebooks-Magics](#) im Verwaltungshandbuch für Amazon EMR.

```
%%configure -f
{
"conf":{

"spark.sql.extensions":"org.apache.iceberg.spark.extensions.IcebergSparkSessionExtensions",
  "spark.sql.catalog.dev":"org.apache.iceberg.spark.SparkCatalog",
  "spark.sql.catalog.dev.type":"hadoop",
```

```
"spark.sql.catalog.dev.warehouse": "s3://DOC-EXAMPLE-BUCKET/example-prefix/"
}
}
```

## Schreiben Sie an einen Iceberg-Tabelle

Das folgende Beispiel zeigt, wie ein Iceberg-Dataset erstellt DataFrame und als Iceberg-Datensatz geschrieben wird. Die Beispiele in diesem Abschnitt veranschaulichen das Arbeiten mit Datensätzen unter Verwendung der Spark-Shell, während eine Verbindung mit dem Hauptknoten mittels SSH als Standard-Benutzer vorhanden ist.

### Note

Um Codebeispiele in die Spark-Shell einzufügen, geben Sie an der Eingabeaufforderung `:paste` ein, fügen das Beispiel ein und drücken dann CTRL+D.

## PySpark

Spark umfasst auch eine auf Python basierende Shell namens `pyspark`, mit der Sie Prototypen von in Python geschriebenen Spark-Programmen entwickeln können. `Invoke` wird in `pyspark` auf dem Hauptknoten festgelegt.

```
## Create a DataFrame.
data = spark.createDataFrame([
    ("100", "2015-01-01", "2015-01-01T13:51:39.340396Z"),
    ("101", "2015-01-01", "2015-01-01T12:14:58.597216Z"),
    ("102", "2015-01-01", "2015-01-01T13:51:40.417052Z"),
    ("103", "2015-01-01", "2015-01-01T13:51:40.519832Z")
],["id", "creation_date", "last_update_time"])

## Write a DataFrame as a Iceberg dataset to the Amazon S3 location.
spark.sql("""CREATE TABLE IF NOT EXISTS dev.db.iceberg_table (id string,
creation_date string,
last_update_time string)
USING iceberg
location 's3://DOC-EXAMPLE-BUCKET/example-prefix/db/iceberg_table'""")

data.writeTo("dev.db.iceberg_table").append()
```

## Scala

```
import org.apache.spark.sql.SaveMode
import org.apache.spark.sql.functions._

// Create a DataFrame.
val data = Seq(
  ("100", "2015-01-01", "2015-01-01T13:51:39.340396Z"),
  ("101", "2015-01-01", "2015-01-01T12:14:58.597216Z"),
  ("102", "2015-01-01", "2015-01-01T13:51:40.417052Z"),
  ("103", "2015-01-01", "2015-01-01T13:51:40.519832Z")
).toDF("id", "creation_date", "last_update_time")

// Write a DataFrame as a Iceberg dataset to the Amazon S3 location.
spark.sql("""CREATE TABLE IF NOT EXISTS dev.db.iceberg_table (id string,
creation_date string,
last_update_time string)
USING iceberg
location 's3://DOC-EXAMPLE-BUCKET/example-prefix/db/iceberg_table'""")

data.writeTo("dev.db.iceberg_table").append()
```

## Lesen von einer Iceberg-Tabelle

### PySpark

```
df = spark.read.format("iceberg").load("dev.db.iceberg_table")
df.show()
```

### Scala

```
val df = spark.read.format("iceberg").load("dev.db.iceberg_table")
df.show()
```

### Spark SQL

```
SELECT * from dev.db.iceberg_table LIMIT 10
```



## Spark-Eigenschaften so konfigurieren, dass der AWS Glue-Datenkatalog als Metastore für Iceberg-Tabellen verwendet wird

Um den AWS Glue-Katalog als Metastore für Iceberg-Tabellen zu verwenden, stellen Sie die Spark-Konfigurationseigenschaften wie folgt ein:

```
spark-submit \  
  --conf spark.sql.catalog.my_catalog=org.apache.iceberg.spark.SparkCatalog \  
  --conf spark.sql.catalog.my_catalog.warehouse=s3://<bucket>/<prefix> \  
  --conf spark.sql.catalog.my_catalog.catalog-  
impl=org.apache.iceberg.aws.glue.GlueCatalog \  
  --conf spark.sql.catalog.my_catalog.io-impl=org.apache.iceberg.aws.s3.S3FileIO \  
  --conf spark.sql.catalog.my_catalog.lock-  
impl=org.apache.iceberg.aws.dynamodb.DynamoDbLockManager \  
  --conf spark.sql.catalog.my_catalog.lock.table=myGlueLockTable
```

## Verwenden Sie einen Iceberg-Cluster mit Trino

Ab Amazon-EMR-Version 6.6.0 können Sie Iceberg mit Ihrem Trino-Cluster verwenden.

In diesem Tutorial verwenden Sie den AWS CLI um mit Iceberg auf einem Amazon EMR Trino-Cluster zu arbeiten. Um die Konsole zur Erstellung eines Clusters mit Iceberg zu verwenden, folgen Sie den Schritten unter [Ein Data Lake von Apache Iceberg mit Amazon Athena, Amazon EMR und AWS Glue erstellen](#).

### Erstellen Sie einen Iceberg-Cluster

Um Iceberg auf Amazon EMR mit dem zu verwenden AWS CLI, erstellen Sie zunächst einen Cluster mit den folgenden Schritten. Informationen zur Spezifizierung der Iceberg-Klassifizierung mithilfe von finden Sie unter oder AWS CLI. [Geben Sie AWS CLI beim Erstellen eines Clusters eine Konfiguration an, indem Sie Beim Erstellen eines Clusters eine Konfiguration mit dem Java SDK angeben](#)

1. Erstellen Sie eine `iceberg.properties`-Datei und legen Sie einen Wert für den ausgewählten Katalog fest. Wenn Sie beispielsweise den Hive-Metastore als Ihren Katalog verwenden möchten, sollte Ihre Datei den folgenden Inhalt haben.

```
connector.name=iceberg  
hive.metastore.uri=thrift://localhost:9083
```

Wenn Sie den AWS Glue-Datenkatalog als Ihren Shop verwenden möchten, sollte Ihre Datei den folgenden Inhalt haben.

```
connector.name=iceberg
iceberg.catalog.type=glue
```

- Erstellen Sie eine Bootstrap-Aktion, die `iceberg.properties` von Amazon S3 nach `/etc/trino/conf/catalog/iceberg.properties` kopiert, wie im folgenden Beispiel. Weitere Informationen zu Bootstrap-Aktionen finden Sie unter [Erstellen von Bootstrap-Aktionen zum Installieren zusätzlicher Software](#).

```
set -ex
sudo aws s3 cp s3://DOC-EXAMPLE-BUCKET/iceberg.properties /etc/trino/conf/catalog/iceberg.properties
```

- Erstellen Sie einen Cluster mit der folgenden Konfiguration und ersetzen Sie dabei den Beispiel-Skriptpfad und den Schlüsselnamen für Bootstrap-Aktionen durch Ihren eigenen.

```
aws emr create-cluster --release-label emr-6.7.0 \
--applications Name=Trino \
--region us-east-1 \
--name My_Trino_Iceberg_Cluster \
--bootstrap-actions '[{"Path":"s3://DOC-EXAMPLE-BUCKET", "Name":"Add
iceberg.properties"}]' \
--instance-groups InstanceGroupType=MASTER,InstanceCount=1,InstanceType=c3.4xlarge
InstanceGroupType=CORE,InstanceCount=3,InstanceType=c3.4xlarge \
--use-default-roles \
--ec2-attributes KeyName=<key-name>
```

## Initialisieren Sie eine Trino-Sitzung für Iceberg

Um eine Trino-Sitzung zu initialisieren, führen Sie den folgenden Befehl aus.

```
trino-cli --catalog iceberg
```

## Schreiben Sie an einen Iceberg-Tabelle

Erstellen Sie Ihre Tabelle und schreiben Sie in Ihre Tabelle mit den folgenden SQL-Befehlen.

```
trino> SHOW SCHEMAS;
trino> CREATE TABLE default.iceberg_table (
    id int,
    data varchar,
    category varchar)
WITH (
    format = 'PARQUET',
    partitioning = ARRAY['category', 'bucket(id, 16)'],
    location = 's3://DOC-EXAMPLE-BUCKET/<prefix>')

trino> INSERT INTO default.iceberg_table VALUES (1,'a','c1'), (2,'b','c2'),
(3,'c','c3');
```

## Lesen von Iceberg aus einer Tabelle

Um von Ihrer Iceberg-Tabelle zu lesen, führen Sie den folgenden Befehl aus.

```
trino> SELECT * from default.iceberg_table;
```

## Verwenden Sie einen Iceberg-Cluster mit Flink

Ab Amazon-EMR-Version 6.9.0 können Sie Iceberg mit einem Flink-Cluster verwenden, ohne dass die Einrichtungsschritte erforderlich sind, die bei der Verwendung der Open-Source-Iceberg-Flink-Integration erforderlich sind.

### Erstellen Sie einen Iceberg-Cluster

Sie können einen Cluster mit installiertem Iceberg mithilfe der AWS Management Console, der AWS CLI oder der Amazon-EMR-API erstellen. In diesem Tutorial verwenden Sie den, AWS CLI um mit Iceberg auf einem Amazon EMR-Cluster zu arbeiten. Um die Konsole zur Erstellung eines Clusters mit Iceberg zu verwenden, folgen Sie den Schritten unter [Ein Data Lake von Apache Iceberg mit Amazon Athena, Amazon EMR und AWS Glue erstellen](#).

Um Iceberg auf Amazon EMR mit dem zu verwenden AWS CLI, erstellen Sie zunächst einen Cluster mit den folgenden Schritten. Informationen zur Spezifizierung der Iceberg-Klassifizierung mithilfe von finden Sie unter oder AWS CLI. [Geben Sie AWS CLI beim Erstellen eines Clusters eine Konfiguration an, indem Sie Beim Erstellen eines Clusters eine Konfiguration mit dem Java SDK angeben](#) Erstellen Sie eine Datei mit dem Namen `configurations.json` und dem folgenden Inhalt:

```
[{
```

```
"Classification":"iceberg-defaults",
  "Properties":{"iceberg.enabled":"true"}
}]
```

Erstellen Sie als Nächstes einen Cluster mit der folgenden Konfiguration und ersetzen Sie den Amazon-S3-Bucket-Beispielpfad und die Subnetz-ID durch Ihre eigenen Werte:

```
aws emr create-cluster --release-label emr-6.9.0 \
--applications Name=Flink \
--configurations file://iceberg_configurations.json \
--region us-east-1 \
--name My_flink_Iceberg_Cluster \
--log-uri s3://DOC-EXAMPLE-BUCKET/ \
--instance-type m5.xlarge \
--instance-count 2 \
--service-role EMR_DefaultRole \
--ec2-attributes InstanceProfile=EMR_EC2_DefaultRole,SubnetId=subnet-1234567890abcdef
```

Alternativ können Sie einen Amazon-EMR-6.9.0-Cluster mit einer Flink-Anwendung darin erstellen und die Datei `/usr/share/aws/iceberg/lib/iceberg-flink-runtime.jar` als JAR-Abhängigkeit in einem Flink-Auftrag verwenden.

## Verwenden des Flink-SQL-Clients

Das SQL-Client-Skript befindet sich unter `/usr/lib/flink/bin`. Sie können das Skript mit dem folgenden Befehl ausführen:

```
flink-yarn-session -d # starting the Flink YARN Session in detached mode
./sql-client.sh
```

Dadurch wird eine Flink-SQL-Shell gestartet.

## Flink-Beispiele

Erstellen Sie eine Iceberg-Tabelle

Flink SQL

```
CREATE CATALOG glue_catalog WITH (
  'type'='iceberg',
  'warehouse'='<WAREHOUSE>',
  'catalog-impl'='org.apache.iceberg.aws.glue.GlueCatalog',
```

```

    'io-impl'='org.apache.iceberg.aws.s3.S3FileIO',
    'lock-impl'='org.apache.iceberg.aws.dynamodb.DynamoDbLockManager',
    'lock.table'='myGlueLockTable'
);

USE CATALOG glue_catalog;

CREATE DATABASE IF NOT EXISTS <DB>;

USE <DB>;

CREATE TABLE IF NOT EXISTS `glue_catalog`.`<DB>`.`sample` (id int, data string);

```

## Tabellen-API

```

EnvironmentSettings settings =
    EnvironmentSettings.newInstance().inBatchMode().build();

TableEnvironment tEnv = TableEnvironment.create(settings);

String warehouse = "<WAREHOUSE>";
String db = "<DB>";

tEnv.executeSql(
    "CREATE CATALOG glue_catalog WITH (\n"
        + "    'type'='iceberg',\n"
        + "    'warehouse'='"
        + warehouse
        + "',\n"
        + "    'catalog-impl'='org.apache.iceberg.aws.glue.GlueCatalog',
\n"
        + "    'io-impl'='org.apache.iceberg.aws.s3.S3FileIO'\n"
        + " );");

tEnv.executeSql("USE CATALOG glue_catalog;");
tEnv.executeSql("CREATE DATABASE IF NOT EXISTS " + db + ";");
tEnv.executeSql("USE " + db + ";");
tEnv.executeSql(
    "CREATE TABLE `glue_catalog`.`" + db + "`.`sample` (id bigint, data string);");

```

## Schreiben Sie an einen Iceberg-Tabelle

### Flink SQL

```
INSERT INTO `glue_catalog`.`<DB>`.`sample` values (1, 'a'),(2,'b'),(3,'c');
```

## Tabellen-API

```
tEnv.executeSql(  
    "INSERT INTO `glue_catalog`.``"  
    + db  
    + "`.`sample` values (1, 'a'),(2,'b'),(3,'c');");
```

## Datenstream-API

```
final StreamExecutionEnvironment env =  
    StreamExecutionEnvironment.getExecutionEnvironment();  
  
final StreamTableEnvironment tableEnv = StreamTableEnvironment.create(env);  
  
String db = "<DB Name>";  
  
String warehouse = "<Warehouse Path>";  
  
GenericRowData rowData1 = new GenericRowData(2);  
rowData1.setField(0, 1L);  
rowData1.setField(1, StringData.fromString("a"));  
  
DataStream<RowData> input = env.fromElements(rowData1);  
  
Map<String, String> props = new HashMap<>();  
props.put("type", "iceberg");  
props.put("warehouse", warehouse);  
props.put("io-impl", "org.apache.iceberg.aws.s3.S3FileIO");  
  
CatalogLoader glueCatalogLoader =  
    CatalogLoader.custom(  
        "glue",  
        props,  
        new Configuration(),  
        "org.apache.iceberg.aws.glue.GlueCatalog");  
  
TableLoader tableLoader =  
    TableLoader.fromCatalog(glueCatalogLoader, TableIdentifier.of(db, "sample"));  
  
DataStreamSink<Void> dataStreamSink =
```

```
FlinkSink.forRowData(input).tableLoader(tableLoader).append();  
  
env.execute("Datastream Write");
```

## Lesen von einer Iceberg-Tabelle

### Flink SQL

```
SELECT * FROM `glue_catalog`.`<DB>`.`sample`;
```

### Tabellen-API

```
Table result = tEnv.sqlQuery("select * from `glue_catalog`.`" + db + "`.`sample`");
```

### Datenstream-API

```
final StreamExecutionEnvironment env =  
    StreamExecutionEnvironment.getExecutionEnvironment();  
  
final StreamTableEnvironment tableEnv = StreamTableEnvironment.create(env);  
  
String db = "<DB Name>";  
  
String warehouse = "<Warehouse Path>";  
  
Map<String, String> props = new HashMap<>();  
props.put("type", "iceberg");  
props.put("warehouse", warehouse);  
props.put("io-impl", "org.apache.iceberg.aws.s3.S3FileIO");  
  
CatalogLoader glueCatalogLoader =  
    CatalogLoader.custom(  
        "glue",  
        props,  
        new Configuration(),  
        "org.apache.iceberg.aws.glue.GlueCatalog");  
  
TableLoader tableLoader =  
    TableLoader.fromCatalog(glueCatalogLoader, TableIdentifier.of(db, "sample"));  
  
DataStream<RowData> batch =  
  
    FlinkSource.forRowData().env(env).tableLoader(tableLoader).streaming(false).build();
```

```
batch.print().name("print-sink");
```

## Verwenden des Hive-Katalogs

Stellen Sie sicher, dass die Abhängigkeiten zwischen Flink und Hive wie unter [Konfigurieren Sie Flink mit Hive Metastore und Glue Catalog](#) beschrieben aufgelöst wurden.

## Flink-Auftrag ausführen

Eine Möglichkeit, einen Auftrag an Flink zu senden, besteht darin, eine Flink-YARN-Sitzung pro Auftrag zu verwenden. Dies kann mit dem folgenden Befehl gestartet werden:

```
sudo flink run -m yarn-cluster -p 4 -yjm 1024m -ytm 4096m $JAR_FILE_NAME
```

## Verwenden Sie einen Iceberg-Cluster mit Hive

Mit Amazon-EMR-Versionen 6.9.0 und höher können Sie Iceberg mit einem Hive-Cluster verwenden, ohne die Einrichtungsschritte durchführen zu müssen, die für Open Source Iceberg Hive Integration erforderlich sind. Für Amazon-EMR-Versionen 6.8.0 und früher können Sie eine Bootstrap-Aktion verwenden, um `iceberg-hive-runtime` JAR für die Konfiguration von Hive für die Iceberg-Unterstützung zu installieren.

Amazon EMR 6.9.0 umfasst alle Features für die [Hive-3.1.3-Integration mit Iceberg 0.14.1](#) und enthält auch zusätzliche Amazon-EMR-Features wie die automatische Auswahl unterstützter Ausführungsgines zur Laufzeit (Amazon EMR auf EKS 6.9.0).

## Erstellen Sie einen Iceberg-Cluster

Sie können einen Cluster mit installiertem Iceberg mithilfe der AWS Management Console, der AWS CLI oder der Amazon EMR-API erstellen. In diesem Tutorial verwenden Sie den, AWS CLI um mit Iceberg auf einem Amazon EMR-Cluster zu arbeiten. Um mit der Konsole einen Cluster mit installiertem Iceberg zu erstellen, folgen Sie den Schritten unter [Ein Data Lake von Apache Iceberg mit Amazon Athena, Amazon EMR und AWS Glue erstellen](#).

Um Iceberg auf Amazon EMR mit dem zu verwenden AWS CLI, erstellen Sie zunächst einen Cluster mit den folgenden Schritten. Informationen zur Spezifizierung der Iceberg-Klassifizierung mithilfe des AWS CLI oder des Java-SDK finden Sie unter oder. [Geben Sie AWS CLI beim Erstellen eines Clusters eine Konfiguration an, indem Sie Beim Erstellen eines Clusters eine Konfiguration mit dem](#)



[Java SDK angeben](#) Erstellen Sie eine Datei mit dem Namen `configurations.json` und dem folgenden Inhalt:

```
[{
  "Classification":"iceberg-defaults",
  "Properties":{"iceberg.enabled":"true"}
}]
```

Erstellen Sie als Nächstes einen Cluster mit der folgenden Konfiguration und ersetzen Sie dabei den Amazon-S3-Bucket-Beispielpfad und die Subnetz-ID durch Ihre eigene:

```
aws emr create-cluster --release-label emr-6.9.0 \
--applications Name=Hive \
--configurations file://iceberg_configurations.json \
--region us-east-1 \
--name My_hive_Iceberg_Cluster \
--log-uri s3://DOC-EXAMPLE-BUCKET/ \
--instance-type m5.xlarge \
--instance-count 2 \
--service-role EMR_DefaultRole \
--ec2-attributes InstanceProfile=EMR_EC2_DefaultRole,SubnetId=subnet-1234567890abcdef
```

Ein Hive-Iceberg-Cluster macht die folgenden Dinge:

- Lädt das Iceberg Hive-Laufzeit-Jar in Hive und aktiviert die Iceberg-bezogene Konfiguration für die Hive-Engine.
- Aktiviert die dynamische Ausführungs-Engine-Auswahl von Amazon EMR Hive, um zu verhindern, dass Benutzer die unterstützte Ausführungs-Engine für die Iceberg-Kompatibilität einrichten.

#### Note

Hive Iceberg-Cluster unterstützen AWS Glue Data Catalog derzeit nicht. Der Standard-Iceberg-Katalog ist `HiveCatalog`, was dem für die Hive-Umgebung konfigurierten Metastore entspricht. Weitere Informationen zur Katalogverwaltung finden Sie unter [Using HCatalog](#) in der [Apache Hive-Dokumentation](#).

## Feature-Unterstützung

Amazon EMR 6.9.0 unterstützt Hive 3.1.3 und Iceberg 0.14.1. Die Featureunterstützung ist auf Iceberg-kompatible Feature für Hive 3.1.2 und 3.1.3 beschränkt. Die folgenden -Befehle sind verfügbar:

- Bei den Amazon-EMR-Versionen 6.9.0 bis 6.12.x müssen Sie das `libfb303-JAR` in das `Hive-auxlib`-Verzeichnis aufnehmen. Verwenden Sie den folgenden Befehl, um es einzubinden:

```
sudo /usr/bin/ln -sf /usr/lib/hive/lib/libfb303-*.jar /usr/lib/hive/auxlib/libfb303.jar
```

Bei Amazon-EMR-Versionen 6.13 und höher wird das `libfb303-JAR` automatisch mit dem `Hive-auxlib`-Verzeichnis verknüpft.

- Erstellen einer Tabelle
  - Nicht partitionierte Tabelle – Externe Tabellen in Hive können erstellt werden, indem der Storage-Handler wie folgt bereitgestellt wird:

```
CREATE EXTERNAL TABLE x (i int) STORED BY  
'org.apache.iceberg.mr.hive.HiveIcebergStorageHandler'
```

- Partitionierte Tabelle – Externe partitionierte Tabellen in Hive können wie folgt erstellt werden:

```
CREATE EXTERNAL TABLE x (i int) PARTITIONED BY (j int) STORED BY  
'org.apache.iceberg.mr.hive.HiveIcebergStorageHandler'
```

### Note

Das `STORED AS`-Dateiformat `ORC/AVRO/PARQUET` wird in Hive 3 nicht unterstützt. Die Standardoption und einzige Option ist Parquet.

- Tabelle löschen – Der `DROP TABLE`-Befehl wird verwendet, um Tabellen zu löschen, wie im folgenden Beispiel:

```
DROP TABLE [IF EXISTS] table_name [PURGE];
```

- Eine Tabelle lesen – `SELECT`-Anweisungen können verwendet werden, um Iceberg-Tabellen in Hive zu lesen, wie im folgenden Beispiel. Unterstützte Ausführungs-Engines sind MR und Tez.

```
SELECT * FROM table_name
```

[Informationen zur Select-Syntax von Hive finden Sie unter Select. LanguageManual](#) Informationen zu Select-Anweisungen mit Iceberg-Tabellen in Hive finden Sie unter [Apache Iceberg Select](#).

- In eine Tabelle einfügen – Die INSERT INTO-HiveQL-Anweisung funktioniert nur für Iceberg-Tabellen mit Unterstützung für die Map-Reduce-Ausführungseingine. Amazon-EMR-Benutzer müssen die Ausführungs-Engine nicht explizit festlegen, da Amazon EMR Hive die Engine für Iceberg Tables zur Laufzeit auswählt.
- Einzelne Tabelle einfügen in – Beispiel:

```
INSERT INTO table_name VALUES ('a', 1);  
INSERT INTO table_name SELECT...;
```

- Einfügen in mehrere Tabellen – Das Einfügen in Anweisungen, die nicht aus mehreren Tabellen bestehen, werden unterstützt. Beispiel:

```
FROM source  
INSERT INTO table_1 SELECT a, b  
INSERT INTO table_2 SELECT c,d;
```

## Überlegungen und Einschränkungen für die Verwendung von Iceberg in Amazon EMR

Dieser Abschnitt enthält Überlegungen und Einschränkungen für die Verwendung von Iceberg mit Spark, Trino, Flink und Hive.

### Überlegungen zur Verwendung von Iceberg mit Spark

- Amazon EMR 6.5.0 unterstützt Iceberg, das auf Amazon EMR auf EKS ausgeführt wird, standardmäßig nicht. Ein benutzerdefiniertes Amazon EMR 6.5.0-Image ist verfügbar, sodass Sie `--jars local:///usr/share/aws/iceberg/lib/iceberg-spark3-runtime.jar` als `spark-submit`-Parameter übergeben können, um Iceberg-Tabellen auf Amazon EMR auf EKS zu erstellen. Weitere Informationen finden Sie unter [Einreichen eines Spark-Workloads in Amazon EMR mithilfe eines benutzerdefinierten Images](#) im Entwicklerhandbuch zu Amazon EMR in EKS.

Sie können sich auch an AWS Support wenden, um Unterstützung zu erhalten. Ab Amazon EMR 6.6.0 wird Iceberg auf Amazon EMR in EKS unterstützt.

- Wenn Sie AWS Glue als Katalog für Iceberg verwenden, stellen Sie sicher, dass die Datenbank, in der Sie eine Tabelle erstellen, in AWS Glue vorhanden ist. Wenn Sie Dienste wie verwenden AWS Lake Formation und den Katalog nicht laden können, stellen Sie sicher, dass Sie über den richtigen Zugriff auf den Dienst verfügen, um den Befehl auszuführen.

## Überlegungen zur Verwendung von Iceberg mit Trino

- Amazon EMR 6.5 bietet keine native Trino-Iceberg-Catalog-Unterstützung für Iceberg. Trino benötigt Iceberg v0.11, daher empfehlen wir, einen Amazon-EMR-Cluster für Trino getrennt vom Spark-Cluster zu starten und Iceberg v0.11 auf diesem Cluster zu integrieren.
- Wenn Sie AWS Glue als Katalog für Iceberg verwenden, stellen Sie sicher, dass die Datenbank, in der Sie eine Tabelle erstellen, in AWS Glue vorhanden ist. Wenn Sie Dienste wie verwenden AWS Lake Formation und den Katalog nicht laden können, stellen Sie sicher, dass Sie über den richtigen Zugriff auf den Dienst verfügen, um den Befehl auszuführen.

## Überlegungen zur Verwendung von Iceberg mit Flink

Wenn Sie AWS Glue als Katalog für Iceberg verwenden, stellen Sie sicher, dass die Datenbank, in der Sie eine Tabelle erstellen, in AWS Glue vorhanden ist. Wenn Sie Dienste wie verwenden AWS Lake Formation und den Katalog nicht laden können, stellen Sie sicher, dass Sie über den richtigen Zugriff auf den Dienst verfügen, um den Befehl auszuführen.

## Überlegungen zur Verwendung von Iceberg mit Hive

- Iceberg unterstützt die folgenden Abfragetypen:
  - Create table
  - Drop table
  - In die Tabelle einfügen
  - Tabelle lesen
- Nur die MR (MapReduce) -Ausführungseingine wird für DML-Operationen (Data Manipulation Language) unterstützt, und MR ist in Hive 3.1.3 veraltet.

- AWS Glue Data Catalog wird derzeit nicht für Iceberg with Hive unterstützt.
- Die Fehlerbehandlung ist nicht robust genug. Im Falle einer Fehlkonfiguration können Einfügungen in Abfragen erfolgreich abgeschlossen werden. Wenn Metadaten nicht aktualisiert werden, kann das jedoch zu Datenverlust führen.

## Versionsverlauf von Iceberg

In der folgenden Tabelle sind die Version von Iceberg aufgeführt, die in jeder Release-Version von Amazon EMR enthalten ist, zusammen mit den Komponenten, die mit der Anwendung installiert wurden. Informationen zu den Komponentenversionen in den einzelnen Versionen finden Sie im Abschnitt [Komponentenversion für Ihre Version in Amazon-EMR-7.x-Versionen](#), [Amazon-EMR-6.x-Versionen](#) oder [Amazon-EMR-5.x-Versionen](#).

### Informationen zur Iceberg-Version

Amazon-EMR-Version sbezeichnung	Iceberg-Version	Mit Iceberg installierte Komponenten
emr-7.1.0	1.4.3-amzn-0	Not available.
emr-7.0.0	1.4.2-amzn-0	Not available.
emr-6.15.0	1.4.0-amzn-0	Not available.
emr-6.14.0	1.3.1-amzn-0	Not available.
emr-6.13.0	1.3.0-amzn-1	Not available.
emr-6.12.0	1.3.0-amzn-0	Not available.
emr-6.11.1	1.2.0-amzn-0	Not available.
emr-6.11.0	1.2.0-amzn-0	Not available.
emr-6.10.1	1.1.0-amzn-0	Not available.
emr-6.10.0	1.1.0-amzn-0	Not available.
emr-6.9.1	0.14.1-amzn-0	Not available.

Amazon-EMR-Version sbezeichnung	Iceberg-Version	Mit Iceberg installierte Komponenten
emr-6.9.0	0.14.1-amzn-0	Not available.
emr-6.8.1	0.14.0-amzn-0	Not available.
emr-6.8.0	0.14.0-amzn-0	Not available.
emr-6.7.0	0.13.1-amzn-0	Not available.
emr-6.6.0	0.13.1	Not available.
emr-6.5.0	0.12.0	Not available.

## Iceberg-Versionshinweise nach Version

- [Amazon EMR 6.9.0 – Iceberg-Versionshinweise](#)

### Amazon EMR 6.9.0 – Iceberg-Versionshinweise

#### Amazon EMR 6.9.0 – Iceberg-Änderungen

Typ	Beschreibung
Funktion	Amazon-EMR-Flink-Integration mit Iceberg.
Funktion	Amazon-EMR-Hive-Integration mit Iceberg.
Funktion	Support für das Zwischenspeichern von Iceberg-Metadatendateien auf Amazon FSx für Lustre, um die Abfrageplanungszeit zu verkürzen.
Backport	<a href="#">PR 5050</a> : Flink 1.15: Support-Schreiboptionen in den Inline-Insert-SQL-Kommentaren.

Typ	Beschreibung
Backport	<a href="#">PR 5282</a> : AWS: Behebung von PUT-Wiederholungsfehlern durch Öffnen neuer Datendateistreams.
Backport	<a href="#">PR 5318</a> : Flink 1.15: Überbrückung der Lücke zwischen FlinkSource und IcebergSource (FLIP-27) und es wurde eine Opt-in-Konfiguration hinzugefügt, um die FLIP-27-Quelle in Flink SQL zu verwenden.
Backport	<a href="#">PR 5344</a> : Flink 1.14: Überbrückung der Lücke zwischen FlinkSource und IcebergSource (FLIP-27) und es wurde eine Opt-in-Konfiguration hinzugefügt, um die FLIP-27-Quelle in Flink SQL zu verwenden.
Backport	<a href="#">PR 5393</a> : Flink 1.14, 1.15: Vermeiden Sie die Konvertierung von Iceberg in Flink-Metriken im FLIP-27-Quelllesegerät. MetricContext
Backport	<a href="#">PR 5401</a> : Flink 1.14, 1.15: Fehlende Gruppe in PR #5393 für FLIP-27-Quelllese-Metriken. IcebergSourceReader
Backport	<a href="#">PR 5679</a> : Spark 3.2, 3.3: Die Übertragung der NULL-Zulässigkeit für den Knoten wurde korrigiert. MergeRows
Backport	<a href="#">PR 5860</a> : Spark 3.3: Fix QueryFailure bei der Ausführung RewriteManifestProcedure auf datumpartitionierten Tabellen.
Backport	<a href="#">PR 5880</a> : Spark 3.3: NULL-Zulässigkeit in Projektionen behoben. merge-on-read
Backport	<a href="#">PR 5917</a> : Spark 3.2: NULL-Zulässigkeit in Projektionen korrigieren. merge-on-read

# Jupyter Notebook in Amazon EMR

[Jupyter Notebook](#) ist eine Open-Source-Webanwendung, mit der Sie Dokumente erstellen und freigeben können, die Live-Code, Gleichungen, Visualisierungen und narrativen Text enthalten. Amazon EMR bietet Ihnen drei Möglichkeiten, mit Jupyter Notebooks zu arbeiten:

Themen

- [EMR Studio](#)
- [Amazon EMR Notebook auf Basis von Jupyter Notebook](#)
- [JupyterHub](#)

## EMR Studio

Amazon EMR Studio ist eine webbasierte integrierte Entwicklungsumgebung (IDE) für vollständig verwaltete [Jupyter Notebooks](#), die auf Amazon-EMR-Clustern ausgeführt werden. Sie können ein EMR Studio für Ihr Team einrichten, um in R, Python, Scala und PySpark geschriebene Anwendungen zu entwickeln, zu visualisieren und zu debuggen.

Wir empfehlen die Verwendung von EMR Studio, wenn Sie Jupyter Notebooks auf Amazon EMR verwenden. Weitere Informationen finden Sie unter [EMR Studio](#) im Verwaltungshandbuch für Amazon EMR.

## Amazon EMR Notebook auf Basis von Jupyter Notebook

EMR Notebooks ist eine in die Amazon-EMR-Konsole integrierte [Jupyter-Notebook-Umgebung](#), mit der Sie schnell Jupyter Notebooks erstellen, sie an Spark-Cluster anhängen und dann den Jupyter-Notebook-Editor in der Konsole öffnen können, um Abfragen und Code remote auszuführen. Ein EMR Notebook wird unabhängig von Clustern in Amazon S3 gespeichert, um dauerhaften Speicher, schnellen Zugriff und Flexibilität zu gewährleisten. Sie können mehrere Notebooks öffnen, mehrere Notebooks an einen einzelnen Cluster anhängen und ein Notizbuch in verschiedenen Clustern wiederverwenden.

Weitere Informationen finden Sie unter [Nutzung von EMR Notebooks](#) im Verwaltungshandbuch für Amazon EMR.



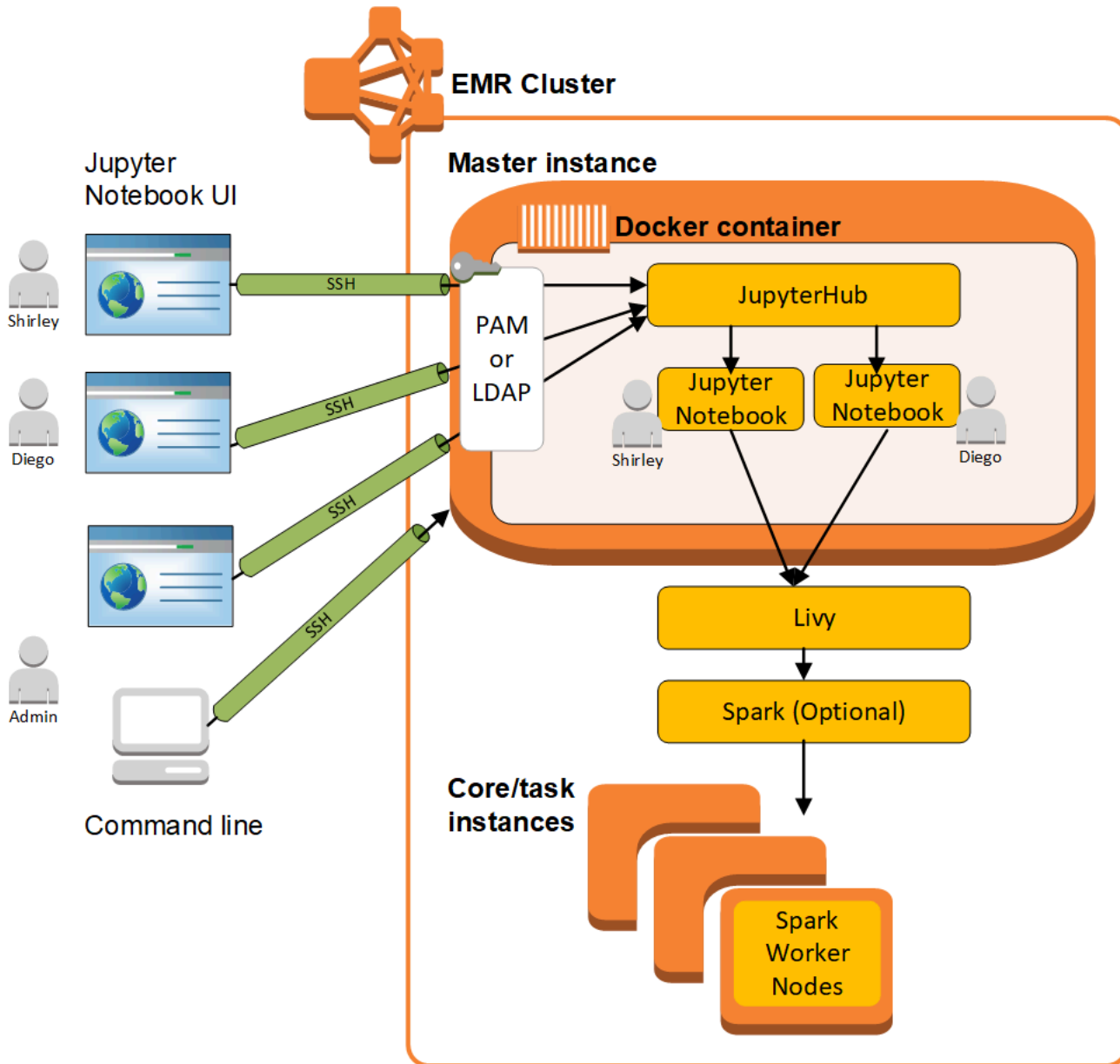
# JupyterHub

[Jupyter Notebook](#) ist eine Open-Source-Webanwendung, mit der Sie Dokumente erstellen und freigeben können, die Live-Code, Gleichungen, Visualisierungen und narrativen Text enthalten.

[JupyterHub](#) ermöglicht es Ihnen, mehrere Instanzen eines Jupyter-Notebookservers für einen einzelnen Benutzer zu hosten. Wenn Sie einen Cluster mit erstellen JupyterHub, erstellt Amazon EMR einen Docker-Container auf dem Master-Knoten des Clusters. JupyterHub, alle für Jupyter und [Sparkmagic](#) erforderlichen Komponenten werden innerhalb des Containers ausgeführt.

Sparkmagic ist eine Kernel-Bibliothek, mit der Jupyter-Notebooks mit [Apache Spark](#) in Amazon EMR über [Apache Livy](#) interagieren. Dabei handelt es sich um einen REST-Server für Spark. Spark und Apache Livy werden automatisch installiert, wenn Sie einen Cluster mit erstellen. JupyterHub Der Standard-Python-3-Kernel für Jupiter ist zusammen mit den Kernen PySpark 3,, und Spark verfügbar PySpark, die mit Sparkmagic verfügbar sind. Sie können diese Kernel verwenden, um ad-hoc-Spark-Code und interaktive SQL-Abfragen mit Python und Scala auszuführen. Sie können zusätzliche Kernel innerhalb des Docker-Containers manuell installieren. Weitere Informationen finden Sie unter [Installieren von zusätzlicher Kernen und Bibliotheken](#).

Das folgende Diagramm zeigt die Komponenten von JupyterHub on Amazon EMR mit den entsprechenden Authentifizierungsmethoden für Notebook-Benutzer und den Administrator. Weitere Informationen finden Sie unter [Hinzufügen von Jupyter-Notebook-Benutzern und Administratoren](#).



In der folgenden Tabelle sind die Version von JupyterHub aufgeführt, die in der neuesten Version der Amazon EMR 7.x-Serie enthalten ist, zusammen mit den Komponenten, mit denen Amazon EMR installiert wird. JupyterHub

Informationen zur Version der Komponenten, mit denen JupyterHub in dieser Version installiert wurde, finden Sie unter Komponentenversionen der [Version 7.1.0](#).

## JupyterHub Versionsinformationen für emr-7.1.0

Amazon-EMR-Version sbezeichnung	JupyterHub Version	Komponenten, die mit installiert wurden JupyterHub
emr-7.1.0	JupyterHub 1.5.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

In der folgenden Tabelle sind die Version von JupyterHub aufgeführt, die in der neuesten Version der Amazon EMR 6.x-Serie enthalten ist, zusammen mit den Komponenten, mit denen Amazon EMR installiert wird. JupyterHub

Informationen zur Version der Komponenten, mit denen JupyterHub in dieser Version installiert wurde, finden Sie unter [Komponentenversionen von Version 6.15.0](#).

## JupyterHub Versionsinformationen für emr-6.15.0

Amazon-EMR-Version sbezeichnung	JupyterHub Version	Komponenten, die mit installiert wurden JupyterHub
emr-6.15.0	JupyterHub 1.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager,

Amazon-EMR-Version sbezeichnung	JupyterHub Version	Komponenten, die mit installiert wurden JupyterHub
		hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

In der folgenden Tabelle sind die Version von JupyterHub aufgeführt, die in der neuesten Version der Amazon EMR 5.x-Serie enthalten ist, zusammen mit den Komponenten, mit denen Amazon EMR installiert wird. JupyterHub

Informationen zur Version der Komponenten, mit denen JupyterHub in dieser Version installiert wurde, finden Sie unter Komponentenversionen von [Version 5.36.2](#).

#### JupyterHub Versionsinformationen für emr-5.36.2

Amazon-EMR-Version sbezeichnung	JupyterHub Version	Komponenten, die mit installiert wurden JupyterHub
emr-5.36.2	JupyterHub 1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Der in Amazon EMR enthaltene JupyterHub Python-3-Kernel ist 3.6.4.

Die im `jupyterhub`-Container installierten Bibliotheken können zwischen Amazon-EMR-Versionen und Amazon-EC2-AMI-Versionen variieren.

Installierte Bibliotheken mit **conda** auflisten

- Führen Sie den folgenden Befehl in der Befehlszeile des Master-Knotens aus:

```
sudo docker exec jupyterhub bash -c "conda list"
```

Installierte Bibliotheken mit **pip** auflisten

- Führen Sie den folgenden Befehl in der Befehlszeile des Master-Knotens aus:

```
sudo docker exec jupyterhub bash -c "pip freeze"
```

Themen

- [Erstellen Sie einen Cluster mit JupyterHub](#)
- [Überlegungen bei der Verwendung JupyterHub auf Amazon EMR](#)
- [Konfiguration JupyterHub](#)
- [Konfigurieren von Persistenz für Notebooks in Amazon S3](#)
- [Eine Verbindung mit dem Hauptknoten und Notebook-Servern herstellen](#)
- [JupyterHub Konfiguration und Verwaltung](#)
- [Hinzufügen von Jupyter-Notebook-Benutzern und Administratoren](#)
- [Installieren von zusätzlicher Kernen und Bibliotheken](#)
- [JupyterHub Versionsverlauf](#)

## Erstellen Sie einen Cluster mit JupyterHub

Sie können einen Amazon EMR-Cluster JupyterHub mithilfe der AWS Management Console, der AWS Command Line Interface, oder der Amazon EMR-API erstellen. Stellen Sie sicher, dass der Cluster nicht mit der Option zum automatischen Beenden nach Abschluss der Schritte angelegt wird (Option `--auto-terminate` in der AWS CLI). Stellen Sie außerdem sicher, dass Administratoren und Notebook-Benutzer auf das Schlüsselpaar zugreifen können, das Sie beim Erstellen des Clusters

verwenden. Weitere Informationen finden Sie unter [Verwenden eines Schlüsselpaars für SSH-Anmeldeinformationen](#) im Verwaltungshandbuch für Amazon EMR.

## Erstellen Sie einen Cluster JupyterHub mithilfe der Konsole

Gehen Sie wie folgt vor, um einen Cluster mit JupyterHub installierten erweiterten Optionen in der Amazon EMR-Konsole zu erstellen.

Um einen Amazon EMR-Cluster zu erstellen, der über die Amazon EMR-Konsole JupyterHub installiert ist

1. Navigieren Sie zur neuen Amazon-EMR-Konsole und wählen Sie in der Seitennavigation die Option Zur alten Konsole wechseln aus. Weitere Informationen darüber, was Sie erwartet, wenn Sie zur alten Konsole wechseln, finden Sie unter [Verwenden der alten Konsole](#).
2. Wählen Sie Create Cluster (Cluster erstellen) und Go to advanced options (Zu erweiterten Optionen) aus.
3. Unter Software Configuration (Softwarekonfiguration):
  - Wählen Sie für Version emr-5.36.2 aus und wählen Sie. JupyterHub
  - Wenn Sie Spark verwenden, um den AWS Glue-Datenkatalog als Metastore für Spark-SQL zu verwenden, wählen Sie Für Spark-Tabellenmetadaten verwenden aus. Weitere Informationen finden Sie unter [Verwenden Sie den AWS Glue-Datenkatalog als Metastore für Spark SQL](#).
  - Für Edit software settings (Softwareeinstellungen bearbeiten) wählen Sie die Option Enter configuration (Konfiguration auswählen) und geben Werte an, oder wählen Load JSON von S3 (JSON aus S3 laden) und geben eine JSON-Konfigurationsdatei an. Weitere Informationen finden Sie unter [Konfiguration JupyterHub](#).
4. Konfigurieren Sie unter Add steps (optional) (Schritte hinzufügen (optional)) die Schritte, die ausgeführt werden sollen, wenn der Cluster erstellt wird, stellen Sie sicher, dass Auto-terminate cluster after the last step is completed (Cluster automatisch beenden, nachdem der letzte Schritt ausgeführt wurde) nicht ausgewählt ist, und klicken Sie auf Next (Weiter).
5. Wählen Sie die Option Hardware Configuration (Hardwarekonfiguration), Next (Weiter). Weitere Informationen finden Sie unter [Konfigurieren von Cluster-Hardware und Netzwerken](#) im Verwaltungshandbuch für Amazon EMR.
6. Wählen Sie Optionen für General Cluster Settings (Allgemeine Cluster-Einstellungen), Next (Weiter).
7. Wählen Sie Security Options (Sicherheitsoptionen), geben Sie ein Schlüsselpaar an und wählen Sie Create Cluster (Cluster erstellen).

## Erstellen Sie einen Cluster JupyterHub mit AWS CLI

Um einen Cluster mit zu starten JupyterHub, verwenden Sie den `aws emr create-cluster` Befehl und geben Sie für die `--applications` Option Folgendes an `Name=JupyterHub`. Im folgenden Beispiel wird ein JupyterHub Cluster auf Amazon EMR mit zwei EC2-Instances (eine Master- und eine Core-Instance) gestartet. Außerdem ist das Debugging aktiviert, wobei die Protokolle am Amazon-S3-Speicherort gespeichert werden wie in `--log-uri` angegeben. Das angegebene Schlüsselpaar bietet Zugriff auf Amazon-EC2-Instances in dem Cluster.

### Note

Linux-Zeilenfortsetzungszeichen (`\`) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (`^`).

```
aws emr create-cluster --name="MyJupyterHubCluster" --release-label emr-5.36.2 \  
--applications Name=JupyterHub --log-uri s3://MyBucket/MyJupyterClusterLogs \  
--use-default-roles --instance-type m5.xlarge --instance-count 2 --ec2-attributes  
KeyName=MyKeyPair
```

## Überlegungen bei der Verwendung JupyterHub auf Amazon EMR

Beachten Sie bei der Verwendung JupyterHub auf Amazon EMR Folgendes.

### Warning

Benutzer-Notebooks und Dateien werden in dem Dateisystem auf dem Master-Knoten gespeichert. Hierbei handelt es sich um flüchtigen Speicher, der nach der Beendigung des Clusters nicht dauerhaft bestehen bleibt. Wenn ein Cluster beendet wird, gehen diese Daten verloren, wenn sie nicht gesichert sind. Wir empfehlen, dass Sie regelmäßige Sicherungen mithilfe von `cron`-Jobs oder anderen für Ihre Anwendung geeigneten Methoden einplanen.

Außerdem bleiben Konfigurationsänderungen innerhalb des Containers bei einem Neustart des Containers möglicherweise nicht bestehen. Wir empfehlen Ihnen, die Containerkonfiguration per Skript vorzunehmen oder anderweitig zu automatisieren, damit Sie Anpassungen leichter reproduzieren können.

- Kerberos-Authentifizierung, die mit einer Amazon-EMR-Sicherheitskonfiguration eingerichtet wurde, wird nicht unterstützt.
- [OAuthenticator](#) wird nicht unterstützt.

## Konfiguration JupyterHub

Sie können die Konfiguration von Notebooks JupyterHub auf Amazon EMR und einzelnen Benutzern anpassen, indem Sie eine Verbindung zum Cluster-Masterknoten herstellen und die Konfigurationsdateien bearbeiten. Nachdem Sie Werte geändert haben, starten Sie den `jupyterhub`-Container.

Ändern Sie die Eigenschaften in den folgenden Dateien, um einzelne JupyterHub Jupyter-Notebooks zu konfigurieren:

- `jupyterhub_config.py` – Diese Datei befindet sich standardmäßig im `/etc/jupyter/conf/`-Verzeichnis auf dem Hauptknoten. Weitere Informationen finden Sie in der Dokumentation unter [Grundlagen der Konfiguration](#). JupyterHub
- `jupyter_notebook_config.py` – Diese Datei wird im `/etc/jupyter/`-Verzeichnis gespeichert und standardmäßig in den `jupyterhub`-Container kopiert. Weitere Informationen finden Sie unter [Konfigurationsdatei und Befehlszeilen-Optionen](#) in der Jupyter-Notebook-Dokumentation.

Sie können auch die Konfigurationsklassifizierung `jupyter-sparkmagic-conf` verwenden, um Sparkmagic anzupassen. Dadurch werden Werte in der `config.json`-Datei für Sparkmagic aktualisiert. Weitere Informationen zu verfügbaren Einstellungen finden Sie unter [example\\_config.json](#) unter [GitHub](#). Weitere Informationen zur Verwendung von Konfigurationsklassifizierungen für Anwendungen in Amazon EMR finden Sie unter [Anwendungen konfigurieren](#).

Im folgenden Beispiel wird ein Cluster mit dem gestartet AWS CLI, der auf die Datei `MyJupyterConfig.json` für die Sparkmagic-Konfigurationsklassifizierungseinstellungen verweist.

### Note

Linux-Zeilenfortsetzungszeichen (`\`) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (`^`).



```
aws emr create-cluster --use-default-roles --release-label emr-5.14.0 \
--applications Name=Jupyter --instance-type m4.xlarge --instance-count 3 \
--ec2-attributes KeyName=MyKey,SubnetId=subnet-1234a5b6 --configurations file://
MyJupyterConfig.json
```

Beispiele für den Inhalt der Datei MyJupyterConfig.json:

```
[
  {
    "Classification": "jupyter-sparkmagic-conf",
    "Properties": {
      "kernel_python_credentials" : "{\"username\":\"diego\",\"base64_password\":
\"mypass\",\"url\":\"http://localhost:8998\",\"auth\":\"None\"}"
    }
  }
]
```

#### Note

Ab Amazon-EMR-Version 5.21.0 können Sie Cluster-Konfigurationen überschreiben und zusätzliche Konfigurationsklassifikationen für jede Instance-Gruppe in einem ausgeführten Cluster angeben. Dazu verwenden Sie die Amazon EMR-Konsole, das AWS Command Line Interface (AWS CLI) oder das AWS SDK. Weitere Informationen finden Sie unter [Angeben einer Konfiguration für eine Instance-Gruppe in einem aktiven Cluster](#).

## Konfigurieren von Persistenz für Notebooks in Amazon S3

Sie können einen JupyterHub Cluster in Amazon EMR so konfigurieren, dass von einem Benutzer gespeicherte Notizbücher in Amazon S3, außerhalb des kurzlebigen Speichers auf Cluster-EC2-Instances, bestehen bleiben.

Sie geben die Amazon-S3-Persistenz mithilfe der `jupyter-s3-conf`-Konfigurationsklassifizierung an, wenn Sie einen Cluster erstellen. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Zusätzlich zur Aktivierung der Amazon-S3-Persistenz mithilfe der Eigenschaft `s3.persistence.enabled` geben Sie einen Bucket in Amazon S3 an, in dem Notebooks mithilfe der Eigenschaft `s3.persistence.bucket` gespeichert werden. Notebooks für jeden Benutzer

werden in einem `jupyter/jupyterhub-user-name`-Ordner im angegebenen Bucket gespeichert. Der Bucket muss bereits in Amazon S3 vorhanden sein und die Rolle für das EC2-Instance-Profil, die Sie bei der Erstellung des Clusters angeben, muss über Berechtigungen für den Bucket verfügen (standardmäßig ist die Rolle `EMR_EC2_DefaultRole`). Weitere Informationen finden [Sie unter Konfigurieren von IAM-Rollen für Amazon EMR-Berechtigungen](#) für Dienste. AWS

Wenn Sie einen neuen Cluster mit denselben Konfigurationsklassifizierungseigenschaften verwenden, können Benutzer Notebooks mit den Inhalten des Speicherorts öffnen.

Beachten Sie, dass wenn Sie Dateien als Module in ein Notebook importieren und Amazon S3 aktiviert haben, dies dazu führt, dass die Dateien auf Amazon S3 hochgeladen werden. Wenn Sie Dateien importieren, ohne die Amazon S3 S3-Persistenz zu aktivieren, werden sie in Ihren JupyterHub Container hochgeladen.

Im folgenden Beispiel wird Amazon-S3-Persistenz aktiviert. Von Benutzern gespeicherte Notebooks werden im Ordner `s3://MyJupyterBackups/jupyter/jupyterhub-user-name` der einzelnen Benutzer gespeichert. Dabei ist *jupyterhub-user-name* ein Benutzername wie beispielsweise `diego`.

```
[
  {
    "Classification": "jupyter-s3-conf",
    "Properties": {
      "s3.persistence.enabled": "true",
      "s3.persistence.bucket": "MyJupyterBackups"
    }
  }
]
```

## Eine Verbindung mit dem Hauptknoten und Notebook-Servern herstellen

JupyterHub Administratoren und Notebook-Benutzer müssen über einen SSH-Tunnel eine Verbindung zum Cluster-Masterknoten herstellen und dann eine Verbindung zu den Webschnittstellen herstellen, die JupyterHub auf dem Master-Knoten bedient werden. Weitere Informationen zum Konfigurieren eines SSH-Tunnels und zur Verwendung des Tunnels als Proxy für Web-Verbindungen finden Sie unter [Mit dem Cluster verbinden](#) im Verwaltungshandbuch für Amazon EMR.

Standardmäßig ist EMR JupyterHub auf Amazon über Port 9443 auf dem Master-Knoten verfügbar. Der interne JupyterHub Proxy bedient auch Notebook-Instances über Port 9443. JupyterHub und auf Jupyter-Webschnittstellen kann über eine URL mit dem folgenden Muster zugegriffen werden:

***https://DNS:9443 MasterNode***

Sie können unter Verwendung der `c.JupyterHub.port`-Eigenschaft in der `jupyterhub_config.py`-Datei einen anderen Port angeben. Weitere Informationen finden Sie in der Dokumentation unter [Netzwerkgrundlagen](#). JupyterHub

Standardmäßig verwendet EMR JupyterHub auf Amazon ein selbstsigniertes Zertifikat für die SSL-Verschlüsselung mit HTTPS. Benutzer werden aufgefordert, das selbstsignierte Zertifikat als vertrauenswürdig einzustufen, wenn sie eine Verbindung herstellen. Sie können ein vertrauenswürdiges Zertifikat und eigene Schlüssel verwenden. Ersetzen Sie die Standard-Zertifikatsdatei, `server.crt`, und die Schlüsseldatei, `server.key`, im `/etc/jupyter/conf/`-Verzeichnis auf dem Master-Knoten durch Ihre eigenen Zertifikat- und Schlüsseldateien. Verwenden Sie die Eigenschaften `c.JupyterHub.ssl_key` und `c.JupyterHub.ssl_cert` der `jupyterhub_config.py`-Datei zur Angabe Ihres SSL-Materialien. Weitere Informationen finden Sie in der [Dokumentation unter Sicherheitseinstellungen](#). JupyterHub Nachdem Sie `jupyterhub_config.py` geändert haben, starten Sie den Container neu.

## JupyterHub Konfiguration und Verwaltung

JupyterHub und verwandte Komponenten laufen in einem Docker-Container mit dem Namen `jupyterhub`, auf dem das Ubuntu-Betriebssystem ausgeführt wird. Sie haben mehrere Möglichkeiten, Komponenten zu verwalten, die innerhalb des Containers ausgeführt werden.

### Warning

Anpassungen, die Sie innerhalb des Containers vornehmen, bleiben möglicherweise nicht erhalten, wenn der Container neu gestartet wird. Wir empfehlen Ihnen, die Containerkonfiguration per Skript vorzunehmen oder anderweitig zu automatisieren, damit Sie Anpassungen leichter reproduzieren können.

## Administration über die Befehlszeile

Wenn Sie mit dem Master-Knoten über SSH verbunden sind, können Sie Befehle über die Docker-Befehlszeilenschnittstelle (CLI) ausgeben und den Container per Namen (`jupyterhub`) oder ID

angeben. Beispiel: `sudo docker exec jupyterhub command` führt Befehle aus, die vom Betriebssystem oder einer im Container ausgeführten Anwendung erkannt werden. Mit dieser Methode können Sie Benutzer zum Betriebssystem hinzufügen und zusätzliche Anwendungen und Bibliotheken innerhalb des Docker-Containers installieren. Beispiel: Das Standard-Container-Image enthält Conda für die Paketinstallation, sodass Sie den folgenden Befehl in der Befehlszeile des Master-Knotens ausführen können, um eine Anwendung, Keras, innerhalb des Containers zu installieren:

```
sudo docker exec jupyterhub conda install keras
```

## Administration durch Übermitteln von Schritten

Schritte sind eine Möglichkeit, Arbeit an einen Cluster zu übermitteln. Schritte können Sie senden, wenn der Cluster gestartet oder ausgeführt wird. Befehle, die Sie in der Befehlszeile ausführen, können mit `command-runner.jar` als Schritte übermittelt werden. Weitere Informationen finden Sie unter [Arbeiten mit Schritten unter Verwendung der CLI und der Konsole](#) im Verwaltungshandbuch für Amazon EMR und [Befehle und Skripte auf einem Amazon-EMR-Cluster ausführen](#).

Sie könnten beispielsweise den folgenden AWS CLI Befehl auf einem lokalen Computer verwenden, um Keras auf die gleiche Weise zu installieren, wie Sie es im vorherigen Beispiel über die Befehlszeile des Master-Knotens getan haben:

```
aws emr add-steps --cluster-id MyClusterID --steps Name="Command Runner",Jar="command-runner.jar",Args="/usr/bin/sudo","/usr/bin/docker","exec","jupyterhub","conda","install","keras"
```

Sie können auch eine Folge von Schritten per Skript festlegen, das Skript in Amazon S3 hochladen und dann `script-runner.jar` verwenden, um das Skript auszuführen, wenn Sie den Cluster erstellen, oder das Skript als Schritt hinzufügen. Weitere Informationen finden Sie unter [Befehle und Skripte auf einem Amazon-EMR-Cluster ausführen](#). Ein Beispiel finden Sie unter [the section called "Beispiel: Bash-Skript zum Hinzufügen mehrerer Benutzer"](#).

## Administration mithilfe von REST-APIs

Jupyter und der HTTP-Proxy JupyterHub stellen REST-APIs bereit, mit denen Sie Anfragen senden können. JupyterHub Um Anfragen an zu senden JupyterHub, müssen Sie der Anfrage ein API-Token beifügen. Mit dem Befehl `curl` können Sie von der Befehlszeile des Master-Knoten aus REST-Befehle ausführen. Weitere Informationen finden Sie in den folgenden Ressourcen:

- [Verwenden JupyterHub der REST-API](#) in der Dokumentation für JupyterHub, die Anweisungen zum Generieren von API-Token enthält
- [Jupyter Notebook-Server-API](#) aktiviert GitHub
- [configurable-http-proxy](#) auf GitHub

Das folgende Beispiel zeigt die Verwendung der REST-API JupyterHub zum Abrufen einer Benutzerliste. Der Befehl übergibt ein zuvor generiertes Admin-Token und verwendet den Standardport 9443 für JupyterHub, um die Ausgabe zur besseren Anzeige an [jq weiterzuleiten](#):

```
curl -XGET -s -k https://$HOST:9443/hub/api/users \  
-H "Authorization: token $admin_token" | jq .
```

## Hinzufügen von Jupyter-Notebook-Benutzern und Administratoren

Sie können eine von zwei Methoden verwenden, um sich für Benutzer zu authentifizieren, JupyterHub sodass sie Notizbücher erstellen und optional verwalten können. JupyterHub Die einfachste Methode ist die Verwendung JupyterHub des Pluggable Authentication Module (PAM). Darüber hinaus unterstützt EMR JupyterHub auf Amazon das [LDAP-Authentifikator-Plugin JupyterHub zum](#) Abrufen von Benutzeridentitäten von einem LDAP-Server, z. B. einem Microsoft Active Directory-Server. Anweisungen und Beispiele für das Hinzufügen von Benutzern mit den beiden Authentifizierungsmethoden finden Sie in diesem Abschnitt.

JupyterHub hat auf Amazon EMR einen Standardbenutzer mit Administratorrechten. Der Benutzername ist  `jovyan` , das Passwort ist  `jupyter` . Wir empfehlen Ihnen dringend, den Benutzer durch einen anderen Benutzer mit Administratorrechten zu ersetzen. Dies ist mit einem Schritt beim Erstellen des Clusters oder durch das Herstellen einer Verbindung zum Master-Knoten möglich, während der Cluster ausgeführt wird.

### Themen

- [Verwenden der PAM-Authentifizierung](#)
- [Verwenden der LDAP-Authentifizierung](#)
- [Benutzer-Identitätswechsel](#)

## Verwenden der PAM-Authentifizierung

Das Erstellen von PAM-Benutzern in JupyterHub Amazon EMR erfolgt in zwei Schritten. Der erste Schritt besteht darin, dem Betriebssystem, das im Container `jupyterhub` auf dem Master-Knoten ausgeführt wird, Benutzer hinzuzufügen, und für jeden Benutzer ein entsprechendes Benutzer-Stammverzeichnis hinzuzufügen. Der zweite Schritt besteht darin, diese Betriebssystembenutzer als JupyterHub Benutzer hinzuzufügen — ein Vorgang, der als Whitelisting bezeichnet wird. JupyterHub Nachdem ein JupyterHub Benutzer hinzugefügt wurde, kann er eine Verbindung zur JupyterHub URL herstellen und seine Betriebssystemanmeldeinformationen für den Zugriff angeben.

Wenn sich ein Benutzer anmeldet, JupyterHub wird die Notebook-Server-Instanz für diesen Benutzer geöffnet, die im Home-Verzeichnis des Benutzers auf dem Master-Knoten gespeichert ist `/var/lib/jupyter/home/username`. Wenn eine Notebook-Server-Instanz nicht existiert, JupyterHub wird eine Notebook-Instanz im Home-Verzeichnis des Benutzers erzeugt. In den folgenden Abschnitten wird gezeigt, wie Benutzer einzeln zum Betriebssystem und zu hinzugefügt werden JupyterHub, gefolgt von einem rudimentären Bash-Skript, das mehrere Benutzer hinzufügt.

### Hinzufügen eines Betriebssystembenutzers zum Container

Das folgende Beispiel verwendet den Befehl [useradd](#) innerhalb des Containers, um einen einzelnen Benutzer hinzuzufügen, `diego`, und ein Stammverzeichnis für diesen Benutzer anzulegen. Der zweite Befehl verwendet [chpasswd](#), um `diego` ein Passwort für diesen Benutzer zu erteilen. Die Befehle werden in der Befehlszeile auf dem Master-Knoten ausgeführt, während mit SSH eine Verbindung eingerichtet wird. Sie können diese Befehle auch mithilfe eines Schritts ausführen, wie zuvor in [Administration durch Übermitteln von Schritten](#) beschrieben.

```
sudo docker exec jupyterhub useradd -m -s /bin/bash -N diego
sudo docker exec jupyterhub bash -c "echo diego:diego | chpasswd"
```

### Einen Benutzer hinzufügen JupyterHub

Sie können das Admin-Panel in JupyterHub oder die REST-API verwenden, um Benutzer und Administratoren oder einfach nur Benutzer hinzuzufügen.

Um Benutzer und Administratoren über das Admin-Panel in hinzuzufügen JupyterHub

1. Stellen Sie über SSH Connect zum Master-Knoten her und melden Sie sich bei `https://MasterNodeDNS:9443` mit einer Identität an, die über Administratorrechte verfügt.
2. Wählen Sie Control Panel (Systemsteuerung), Admin (Administrator).

3. Wählen Sie User (Benutzer), Add Users (Benutzer hinzufügen) oder wählen Sie Admin (Administrator), Add Admins (Administratoren hinzufügen).

So fügen Sie einen Benutzer mit dem REST API hinzu

1. Stellen Sie unter Verwendung von SSH eine Verbindung mit dem Master-Knoten her und verwenden Sie den folgenden Befehl auf dem Master-Knoten, oder führen Sie den Befehl als Schritt aus.
2. Erwerben Sie ein Administrator-Token, um API-Anfragen zu stellen, und ersetzen Sie es *AdminToken* im folgenden Schritt durch dieses Token.
3. Verwenden Sie den folgenden Befehl und *UserName* ersetzen Sie ihn durch einen Betriebssystembenutzer, der im Container erstellt wurde.

```
curl -XPOST -H "Authorization: token AdminToken" "https://$(hostname):9443/hub/api/users/UserName"
```

#### Note

Sie werden automatisch als Benutzer JupyterHub ohne Administratorrechte hinzugefügt, wenn Sie sich zum ersten Mal an der JupyterHub Weboberfläche anmelden.

Beispiel: Bash-Skript zum Hinzufügen mehrerer Benutzer

Das folgende Bash-Beispielskript verknüpft die vorherigen Schritte in diesem Abschnitt, um mehrere JupyterHub Benutzer zu erstellen. Das Skript kann direkt auf dem Hauptknoten ausgeführt werden, oder in Amazon S3 hochgeladen und als Schritt ausgeführt werden.

Das Skript stellt zuerst ein Array mit Benutzernamen, und verwendet den Befehl `jupyterhub token` zum Erstellen eines API-Token für den Standard-Administrator, `joyan`. Anschließend erstellt es einen Betriebssystem-Benutzer im `jupyterhub`-Container für jeden Benutzer und weist jedem von ihnen ein anfängliches Passwort zu, das gleich ihrem Benutzernamen ist. Schließlich ruft es die REST-API-Operation auf, um jeden Benutzer in JupyterHub zu erstellen. Es übergibt das Token, das zuvor im Skript erstellt wurde und gibt die REST-Antwort der besseren Übersicht halber per Pipe an `jq` weiter.

```
# Bulk add users to container and JupyterHub with temp password of username
```

```

set -x
USERS=(shirley diego ana richard li john mary anaya)
TOKEN=$(sudo docker exec jupyterhub /opt/conda/bin/jupyterhub token jovyan | tail -1)
for i in "${USERS[@]}";
do
    sudo docker exec jupyterhub useradd -m -s /bin/bash -N $i
    sudo docker exec jupyterhub bash -c "echo $i:$i | chpasswd"
    curl -XPOST --silent -k https://$(hostname):9443/hub/api/users/$i \
-H "Authorization: token $TOKEN" | jq
done

```

Speichern Sie das Skript an einem Speicherort in Amazon S3, beispielsweise `s3://mybucket/createjupyterusers.sh`. Anschließend können Sie es mit `script-runner.jar` als Schritt ausführen.

Beispiel: Ausführen des Skripts beim Erstellen eines Clusters (AWS CLI)

#### Note

Linux-Zeilenfortsetzungszeichen (`\`) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (`^`).

```

aws emr create-cluster --name="MyJupyterHubCluster" --release-label emr-5.36.2 \
--applications Name=JupyterHub --log-uri s3://MyBucket/MyJupyterClusterLogs \
--use-default-roles --instance-type m5.xlarge --instance-count 2 --ec2-attributes
KeyName=MyKeyPair \
--steps Type=CUSTOM_JAR,Name=CustomJAR,ActionOnFailure=CONTINUE,\
Jar=s3://region.elasticmapreduce/libs/script-runner/script-runner.jar,Args=["s3://
mybucket/createjupyterusers.sh"]

```

Ausführen des Skripts auf einem vorhandenen Cluster (AWS CLI)

#### Note

Linux-Zeilenfortsetzungszeichen (`\`) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (`^`).



```
aws emr add-steps --cluster-id j-XXXXXXXX --steps Type=CUSTOM_JAR,\
Name=CustomJAR,ActionOnFailure=CONTINUE,\
Jar=s3://region.elasticmapreduce/libs/script-runner/script-runner.jar,Args=["s3://\
mybucket/createjupyterusers.sh"]
```

## Verwenden der LDAP-Authentifizierung

Lightweight Directory Access Protocol (LDAP) ist ein Anwendungsprotokoll zum Abfragen und Ändern von Objekten, die Ressourcen wie Benutzern und Computern entsprechen, die bei einem LDAP-kompatiblen Verzeichnisdiensteanbieter wie Active Directory oder auf einem OpenLDAP-Server gespeichert sind. Sie können das [LDAP-Authentifikator-Plug-In für JupyterHub](#) with JupyterHub auf Amazon EMR verwenden, um LDAP für die Benutzerauthentifizierung zu verwenden. Das Plugin verwaltet Anmeldesitzungen für LDAP-Benutzer und stellt Jupyter Benutzerinformationen zur Verfügung. Auf diese Weise können Benutzer mithilfe der auf einem LDAP-kompatiblen Server gespeicherten Anmeldeinformationen für ihre Identitäten eine Verbindung zu JupyterHub und Notebooks herstellen.

Die Schritte in diesem Abschnitt führen Sie durch die folgenden Schritte zur Einrichtung und Aktivierung von LDAP mithilfe des LDAP Authenticator-Plug-ins für JupyterHub. Sie führen die Schritte aus, während Sie mit der Befehlszeile des Master-Knotens verbunden sind. Weitere Informationen finden Sie unter [Eine Verbindung mit dem Hauptknoten und Notebook-Servern herstellen](#).

1. Erstellen Sie eine LDAP-Konfigurationsdatei mit Informationen über den LDAP-Server, wie Host-IP-Adresse, Port, Bindungsnamen usw.
2. Ändern Sie `/etc/jupyter/conf/jupyterhub_config.py`, um das LDAP Authenticator Plugin für zu aktivieren. JupyterHub
3. Erstellen Sie ein Skript, das LDAP innerhalb des `jupyterhub`-Containers konfiguriert, und führen Sie es aus.
4. Fragen Sie LDAP nach Benutzern ab und erstellen Sie dann Basisverzeichnisse innerhalb des Containers für jeden Benutzer. JupyterHub benötigt Home-Verzeichnisse zum Hosten von Notebooks.
5. Führen Sie ein Skript aus, das neu gestartet wird JupyterHub

**⚠ Important**

Bevor Sie LDAP einrichten, testen Sie Ihre Netzwerkinfrastruktur, um sicherzustellen, dass der LDAP-Server und der Cluster-Master-Knoten wie gefordert kommunizieren können. TLS verwendet in der Regel Port 389 über eine einfache TCP-Verbindung. Wenn Ihre LDAP-Verbindung SSL verwendet, ist der bekannte TCP-Port für SSL 636.

## Die LDAP-Konfigurationsdatei erstellen

Das folgende Beispiel verwendet die folgenden Platzhalter-Konfigurationswerte. Ersetzen Sie diese durch Parameter, die Ihrer Implementierung entsprechen.

- Der LDAP-Server läuft in Version 3 und ist auf Port 389 verfügbar. Dies ist der standardmäßige Nicht-SSL-Port für LDAP.
- Der grundlegende spezifische Name (DN, Distinguished Name) ist `dc=example, dc=org`.

Erstellen Sie mit einem Texteditor die Datei [ldap.conf](#) mit Inhalt, der etwa wie folgt aussieht. Verwenden Sie geeignete Werte für Ihre LDAP-Implementierung. Ersetzen Sie `host` durch die IP-Adresse oder den auflösbaren Hostnamen Ihres LDAP-Servers.

```
base dc=example,dc=org
uri ldap://host
ldap_version 3
binddn cn=admin,dc=example,dc=org
bindpw admin
```

## Aktivieren Sie das LDAP Authenticator Plugin für JupyterHub

Verwenden Sie einen Texteditor, um die Datei `/etc/jupyter/conf/jupyterhub_config.py` zu ändern, und fügen Sie [ldapauthenticator](#)-Eigenschaften hinzu, etwa wie folgt. Ersetzen Sie `host` durch die IP-Adresse oder den auflösbaren Hostnamen des LDAP-Servers. Das Beispiel geht davon aus, dass sich die Benutzerobjekte innerhalb einer Organisationseinheit (ou) mit dem Namen `people` befinden, und verwendet die spezifischen Namenskomponenten, die Sie zuvor mit `ldap.conf` eingerichtet haben.

```
c.JupyterHub.authenticator_class = 'ldapauthenticator.LDAPAuthenticator'
c.LDAPAuthenticator.use_ssl = False
```

```
c.LDAPAuthenticator.server_address = 'host'  
c.LDAPAuthenticator.bind_dn_template = 'cn={username},ou=people,dc=example,dc=org'
```

## LDAP im Container konfigurieren

Erstellen Sie mithilfe eines Texteditors ein Bash-Skript mit folgendem Inhalt:

```
#!/bin/bash  
  
# Uncomment the following lines to install LDAP client libraries only if  
# using Amazon EMR release version 5.14.0. Later versions install libraries by default.  
# sudo docker exec jupyterhub bash -c "sudo apt-get update"  
# sudo docker exec jupyterhub bash -c "sudo apt-get -y install libnss-ldap libpam-ldap  
  ldap-utils nscd"  
  
# Copy ldap.conf  
sudo docker cp ldap.conf jupyterhub:/etc/ldap/  
sudo docker exec jupyterhub bash -c "cat /etc/ldap/ldap.conf"  
  
# configure nss switch  
sudo docker exec jupyterhub bash -c "sed -i 's/\(^passwd.*\)/\1 ldap/g' /etc/  
nsswitch.conf"  
sudo docker exec jupyterhub bash -c "sed -i 's/\(^group.*\)/\1 ldap/g' /etc/  
nsswitch.conf"  
sudo docker exec jupyterhub bash -c "sed -i 's/\(^shadow.*\)/\1 ldap/g' /etc/  
nsswitch.conf"  
sudo docker exec jupyterhub bash -c "cat /etc/nsswitch.conf"  
  
# configure PAM to create home directories  
sudo docker exec jupyterhub bash -c "echo 'session required          pam_mkhomedir.so  
  skel=/etc/skel umask=077' >> /etc/pam.d/common-session"  
sudo docker exec jupyterhub bash -c "cat /etc/pam.d/common-session"  
  
# restart nscd service  
sudo docker exec jupyterhub bash -c "sudo service nscd restart"  
  
# Test  
sudo docker exec jupyterhub bash -c "getent passwd"  
  
# Install ldap plugin  
sudo docker exec jupyterhub bash -c "pip install jupyterhub-ldapauthenticator"
```

Speichern Sie das Skript auf dem Master-Knoten und führen Sie es in der Befehlszeile des Master-Knotens aus. Wenn das Skript beispielsweise als `configure_ldap_client.sh` gespeichert ist, machen Sie die Datei ausführbar:

```
chmod +x configure_ldap_client.sh
```

Und führen das Skript aus:

```
./configure_ldap_client.sh
```

## Attribute zur Active Directory hinzufügen

Um jeden Benutzer zu finden und den entsprechenden Eintrag in der Datenbank zu erstellen, benötigt der JupyterHub Docker-Container die folgenden UNIX-Eigenschaften für das entsprechende Benutzerobjekt in Active Directory. Weitere Informationen finden Sie im Abschnitt [Wie kann ich die GUID/UID RFC 2307-Attribute weiter bearbeiten, nachdem das Unix-Attribute-Plug-in nicht mehr für die Active-Directory-Benutzer und Computers-MMC-Snap-in verfügbar ist?](#) im Artikel [Klärung des Status der Identitätsverwaltung für Unix \(IDMU\) und der NIS-Serverrolle in der technischen Vorschau von Windows Server 2016 und darüber hinaus](#).

- `homeDirectory`

Dies ist der Speicherort im Stammverzeichnis des Benutzers, das in der Regel `/home/username` ist.

- `gidNumber`

Dies ist ein Wert größer als 60.000, der noch nicht von einem anderen Benutzer verwendet wird. Überprüfen Sie die `etc/passwd`-Datei auf GIDs, die bereits verwendet werden.

- `uidNumber`

Dies ist ein Wert größer als 60000, der noch nicht von einer anderen Gruppe verwendet wird. Überprüfen Sie die `etc/group`-Datei auf UIDs, die bereits verwendet werden.

- `uid`

Dies ist mit *Benutzername* identisch.

## Root-Verzeichnisse für Benutzer erstellen

JupyterHub benötigt Home-Verzeichnisse innerhalb des Containers, um LDAP-Benutzer zu authentifizieren und Instanzdaten zu speichern. Das folgende Beispiel zeigt zwei Benutzer, shirley und diego, im LDAP-Verzeichnis.

Im ersten Schritt wird der LDAP-Server nach den Benutzer-IDs und Gruppen-IDs der Benutzer abgefragt. Dazu wird [ldapsearch](#) verwendet, wie im nachfolgenden Beispiel gezeigt. Ersetzen Sie *host* durch die IP-Adresse oder den auflösbaren Hostnamen Ihres LDAP-Servers:

```
ldapsearch -x -H ldap://host \  
-D "cn=admin,dc=example,dc=org" \  
-w admin \  
-b "ou=people,dc=example,dc=org" \  
-s sub \  
"(objectclass=*)" uidNumber gidNumber
```

Der Befehl `ldapsearch` gibt eine LDIF-formatierte Antwort zurück, die für die Benutzer shirley und diego etwa wie nachfolgend gezeigt aussieht.

```
# extended LDIF  
  
# LDAPv3  
# base <ou=people,dc=example,dc=org> with scope subtree  
# filter: (objectclass=*)  
# requesting: uidNumber gidNumber sn  
  
# people, example.org  
dn: ou=people,dc=example,dc=org  
  
# diego, people, example.org  
dn: cn=diego,ou=people,dc=example,dc=org  
sn: B  
uidNumber: 1001  
gidNumber: 100  
  
# shirley, people, example.org  
dn: cn=shirley,ou=people,dc=example,dc=org  
sn: A  
uidNumber: 1002  
gidNumber: 100
```

```
# search result
search: 2
result: 0 Success

# numResponses: 4
# numEntries: 3
```

Unter Verwendung der Informationen aus der Antwort führen Sie Befehle innerhalb des Containers zum Erstellen eines Stammverzeichnisses für jeden allgemeinen Namen (cn) des Benutzers aus. Verwenden Sie `uidNumber` und `gidNumber`, um das Eigentum für das Stammverzeichnis für diesen Benutzer festzulegen. Die folgenden Beispielbefehle erledigen dies für den Benutzer *shirley*.

```
sudo docker container exec jupyterhub bash -c "mkdir /home/shirley"
sudo docker container exec jupyterhub bash -c "chown -R $uidNumber /home/shirley"
sudo docker container exec jupyterhub bash -c "sudo chgrp -R $gidNumber /home/shirley"
```

### Note

Der LDAP-Authentifikator für unterstützt die lokale Benutzererstellung JupyterHub nicht. Weitere Informationen finden Sie unter [LDAP-Authentifikator-Konfigurationshinweis zur lokalen Benutzererstellung](#).

Um einen lokalen Benutzer manuell zu erstellen, verwenden Sie den folgenden Befehl.

```
sudo docker exec jupyterhub bash -c "echo 'shirley:x:$uidNumber:$gidNumber::/home/shirley:/bin/bash' >> /etc/passwd"
```

Starten Sie den Container neu JupyterHub

Führen Sie den folgenden Befehl aus, um den `jupyterhub`-Container neu zu starten:

```
sudo docker stop jupyterhub
sudo docker start jupyterhub
```

## Benutzer-Identitätswechsel

Ein Spark-Auftrag, der in einem Jupyter-Notebook ausgeführt wird, durchläuft während seiner Ausführung in Amazon EMR mehrere Anwendungen. Zum Beispiel wird PySpark 3-Code, den ein

Benutzer in Jupyter ausführt, von Sparkmagic empfangen. Sparkmagic verwendet eine HTTP-POST-Anfrage, um ihn an Livy zu senden, das dann einen Spark-Job erstellt, der mit YARN auf dem Cluster ausgeführt wird.

Standardmäßig werden auf diese Weise gesendete YARN-Aufträge als Benutzer `livy` ausgeführt. Dies gilt unabhängig von dem Benutzer, der den Auftrag initiiert hat. Durch das Einrichten eines Benutzer-Identitätswechsels können Sie die Benutzer-ID des Notebook-Benutzers auch für den dem YARN-Auftrag zugeordneten Benutzer verwenden. Von `shirley` und `diego` initiierte Aufträge werden nicht dem Benutzer `livy` zugeordnet. Stattdessen werden die von den Benutzern initiierten Aufträge `shirley` und `diego` zugeordnet. Dies erleichtert das Überwachen der Jupyter-Nutzung und das Verwalten von Anwendungen in der Organisation.

Diese Konfiguration wird nur unterstützt, wenn Aufrufe von Sparkmagic an Livy ohne Authentifizierung erfolgen. Anwendungen, die eine Authentifizierung oder einen Proxy-Layer zwischen Hadoop-Anwendungen und Livy (z. B. Apache Knox Gateway) bereitstellen, werden nicht unterstützt. Bei den Schritten zur Konfiguration des Benutzerwechsels in diesem Abschnitt wird davon ausgegangen, dass JupyterHub und Livy auf demselben Master-Knoten ausgeführt werden. Wenn die Anwendung separate Cluster enthält, muss [Schritt 3: HDFS-Stammverzeichnisse für Benutzer erstellen](#) geändert werden, damit HDFS-Verzeichnisse auf dem Livy-Master-Knoten erstellt werden.

Schritte zum Konfigurieren des Benutzer-Identitätswechsels

- [Schritt 1: Livy konfigurieren](#)
- [Schritt 2: Benutzer hinzufügen](#)
- [Schritt 3: HDFS-Stammverzeichnisse für Benutzer erstellen](#)

Schritt 1: Livy konfigurieren

Sie verwenden die Konfigurationsklassifizierungen `livy-conf` und `core-site`, wenn Sie einen Cluster erstellen, um Livy-Benutzer-Identitätswechsel zu ermöglichen (siehe folgendes Beispiel). Speichern Sie die Konfigurationsklassifizierung als JSON-Objekt und verweisen Sie dann darauf, wenn Sie den Cluster erstellen. Sie können die Konfigurationsklassifizierung auch im Code angeben. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

```
[
  {
    "Classification": "livy-conf",
    "Properties": {
      "livy.impersonation.enabled": "true"
    }
  }
]
```

```

    }
  },
  {
    "Classification": "core-site",
    "Properties": {
      "hadoop.proxyuser.livy.groups": "*",
      "hadoop.proxyuser.livy.hosts": "*"
    }
  }
]

```

## Schritt 2: Benutzer hinzufügen

Fügen Sie JupyterHub Benutzer mithilfe von PAM oder LDAP hinzu. Weitere Informationen finden Sie unter [Verwenden der PAM-Authentifizierung](#) und [Verwenden der LDAP-Authentifizierung](#).

## Schritt 3: HDFS-Stammverzeichnisse für Benutzer erstellen

Sie haben eine Verbindung zum Master-Knoten hergestellt, um Benutzer zu erstellen. Kopieren Sie den unten angegebenen Inhalt, während die Verbindung zum Master-Knoten noch besteht, und speichern Sie ihn in einer Skriptdatei. Das Skript erstellt HDFS-Home-Verzeichnisse für jeden JupyterHub Benutzer auf dem Master-Knoten. Das Skript unterstellt, dass Sie die Benutzer-ID des Standardadministrators (*jovyan*) verwenden.

```

#!/bin/bash

CURL="curl --silent -k"
HOST=$(curl -s http://169.254.169.254/latest/meta-data/local-hostname)

admin_token() {
  local user=jovyan
  local pwd=jupyter
  local token=$(($CURL https://$HOST:9443/hub/api/authorizations/token \
    -d "{\"username\":\"$user\", \"password\":\"$pwd\"}" | jq ".token")
  if [[ $token != null ]]; then
    token=$(echo $token | sed 's//g')
  else
    echo "Unable to get Jupyter API Token."
    exit 1
  fi
  echo $token
}

```



```
# Get Jupyter Admin token
token=$(admin_token)

# Get list of Jupyter users
users=$(curl -XGET -s -k https://$HOST:9443/hub/api/users \
  -H "Authorization: token $token" | jq '.[].name' | sed 's/"//g')

# Create HDFS home dir
for user in ${users[@]};
do
  echo "Create hdfs home dir for $user"
  hadoop fs -mkdir /user/$user
  hadoop fs -chmod 777 /user/$user
done
```

## Installieren von zusätzlicher Kernen und Bibliotheken

Wenn Sie einen Cluster mit JupyterHub Amazon EMR erstellen, werden der Standard-Python-3-Kernel für Jupyter zusammen mit den PySpark und Spark-Kernel für Sparkmagic auf dem Docker-Container installiert. Sie können zusätzliche Kernel installieren. Sie können auch zusätzliche Bibliotheken und Pakete installieren und dann für die entsprechende Shell importieren.

### Installieren eines Kernels

Kernel werden innerhalb des Docker-Containers installiert. Am einfachsten ist dies zu bewerkstelligen, indem Sie ein Bash-Skript mit Installationsbefehlen erstellen, es auf dem Master-Knoten speichern und es dann mit dem Befehl `sudo docker exec jupyterhub script_name` im Container `jupyterhub` ausführen. Das folgende Beispielskript installiert den Kernel und installiert dann einige Bibliotheken für diesen Kernel auf dem Master-Knoten, sodass Sie die Bibliotheken später unter Verwendung des Kernels in Jupyter importieren können.

```
#!/bin/bash

# Install Python 2 kernel
conda create -n py27 python=2.7 anaconda
source /opt/conda/envs/py27/bin/activate
apt-get update
apt-get install -y gcc
/opt/conda/envs/py27/bin/python -m pip install --upgrade ipykernel
/opt/conda/envs/py27/bin/python -m ipykernel install
```

```
# Install libraries for Python 2
/opt/conda/envs/py27/bin/pip install paramiko nltk scipy numpy scikit-learn pandas
```

Um den Kernel und Bibliotheken innerhalb des Containers zu installieren, öffnen Sie eine Terminal-Verbindung zu dem Master-Knoten, speichern das Skript in `/etc/jupyter/install_kernels.sh` und führen den folgenden Befehl in der Master-Knoten Befehlszeile aus:

```
sudo docker exec jupyterhub bash /etc/jupyter/install_kernels.sh
```

## Verwenden von Bibliotheken und Installieren von zusätzlichen Bibliotheken

Ein Kernsatz von Bibliotheken für maschinelles Lernen und Datenwissenschaft für Python 3 ist JupyterHub auf Amazon EMR vorinstalliert. Sie können `sudo docker exec jupyterhub bash -c "conda list"` und `sudo docker exec jupyterhub bash -c "pip freeze"` verwenden.

Wenn ein Spark-Job Bibliotheken auf Worker-Knoten benötigt, empfehlen wir, dass Sie beim Erstellen des Clusters ein Skript ausführen, um die Bibliotheken zu installieren. Bootstrap-Aktionen werden beim Erstellungsprozess des Clusters auf allen Cluster-Knoten ausgeführt, was die Installation vereinfacht. Wenn Sie Bibliotheken auf Core-/Worker-Knoten installieren, nachdem ein Cluster läuft, ist die Operation komplizierter. Wir zeigen in diesem Abschnitt ein Python-Beispielprogramm, das demonstriert, wie diese Bibliotheken zu installieren sind.

Die in diesem Abschnitt gezeigten Beispiele für Bootstrap-Aktionen und Python-Programme verwenden ein Bash-Skript, das in Amazon S3 gespeichert ist, um die Bibliotheken auf allen Knoten zu installieren.

Das in den folgenden Beispielen verwendete Skript `pip` installiert `paramiko`, `nltk`, `scipy`, `scikit-learn` und `pandas` für den Python 3-Kernel:


```
#!/bin/bash

sudo python3 -m pip install boto3 paramiko nltk scipy scikit-learn pandas
```

Nachdem Sie das Skript erstellen, laden Sie es an einem Speicherort in Amazon S3 hoch, z. B. `s3://mybucket/install-my-jupyter-libraries.sh`. Weitere Informationen finden Sie unter [Hochladen von Objekten](#) im Benutzerhandbuch für Amazon Simple Storage Service, sodass Sie es in Ihrer Bootstrap-Aktion oder in Ihrem Python-Programm verwenden können.

Um eine Bootstrap-Aktion anzugeben, mit der Bibliotheken auf allen Knoten installiert werden, wenn Sie einen Cluster mit dem AWS CLI

1. Erstellen Sie ein Skript ähnlich dem früheren Beispiel und speichern Sie es an einem Speicherort in Amazon S3. Wir verwenden das Beispiel `s3://mybucket/install-my-jupyter-libraries.sh`.
2. Erstellen Sie den Cluster mit JupyterHub und verwenden Sie das Path Argument der `--bootstrap-actions` Option, um den Speicherort des Skripts anzugeben, wie im folgenden Beispiel gezeigt:

 Note

Linux-Zeilenumbruchzeichen (`\`) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (`^`).

```
aws emr create-cluster --name="MyJupyterHubCluster" --release-label emr-5.36.2 \  
--applications Name=JupyterHub --log-uri s3://MyBucket/MyJupyterClusterLogs \  
--use-default-roles --instance-type m5.xlarge --instance-count 2 --ec2-attributes \  
KeyName=MyKeyPair \  
--bootstrap-actions Path=s3://mybucket/install-my-jupyter- \  
libraries.sh,Name=InstallJupyterLibs
```

So geben Sie eine Bootstrap-Aktion an, die Bibliotheken auf allen Knoten installiert, wenn Sie einen Cluster mithilfe der Konsole erstellen

1. Navigieren Sie zur neuen Amazon-EMR-Konsole und wählen Sie in der Seitennavigation die Option Zur alten Konsole wechseln aus. Weitere Informationen darüber, was Sie erwartet, wenn Sie zur alten Konsole wechseln, finden Sie unter [Verwenden der alten Konsole](#).
2. Wählen Sie Create Cluster (Cluster erstellen) und Go to advanced options (Zu erweiterten Optionen) aus.
3. Geben Sie Einstellungen für Software and Steps (Software und Schritte) und Hardware an, wie für Ihre Anwendung geeignet.
4. Auf dem Bildschirm General Cluster Settings (Allgemeine Cluster-Einstellungen) expandieren Sie Bootstrap Actions (Bootstrap-Aktionen).

5. Wählen Sie für Add bootstrap action (Bootstrap-Aktion hinzufügen) die Option Custom action (Benutzerdefinierte Aktion) und Configure and add (Konfigurieren und hinzufügen) aus.
6. Geben Sie für Name einen benutzerfreundlichen Namen ein. Geben Sie unter Skriptspeicherort den Speicherort Ihres Skripts in Amazon S3 ein (das Beispiel, das wir verwenden, ist s3://mybucket/install-my-jupyter-libraries .sh). Tragen Sie in Optional arguments (Optionale Argumente) nichts ein und wählen Sie Add (Hinzufügen).
7. Geben Sie andere Einstellungen für Ihren Cluster an und wählen Sie Next (Weiter).
8. Geben Sie Sicherheitseinstellungen an und wählen Sie Create cluster (Cluster erstellen).

### Example Installieren von Bibliotheken auf Core-Knoten eines aktiven Clusters

Nach der Installation von Bibliotheken auf dem Master-Knoten innerhalb von Jupyter können Sie auf verschiedene Arten Bibliotheken auf in Ausführung befindlichen Core-Knoten installieren. Das folgende Beispiel zeigt ein Python-Programm, das für die Ausführung auf einem lokalen Computer geschrieben wurde. Wenn Sie das Python-Programm lokal ausführen, verwendet es das AWS-RunShellScript von, AWS Systems Manager um das weiter oben in diesem Abschnitt gezeigte Beispielskript auszuführen, das Bibliotheken auf den Kernknoten des Clusters installiert.

```
import argparse
import time
import boto3

def install_libraries_on_core_nodes(cluster_id, script_path, emr_client, ssm_client):
    """
    Copies and runs a shell script on the core nodes in the cluster.

    :param cluster_id: The ID of the cluster.
    :param script_path: The path to the script, typically an Amazon S3 object URL.
    :param emr_client: The Boto3 Amazon EMR client.
    :param ssm_client: The Boto3 AWS Systems Manager client.
    """
    core_nodes = emr_client.list_instances(
        ClusterId=cluster_id, InstanceGroupTypes=["CORE"]
    )["Instances"]
    core_instance_ids = [node["Ec2InstanceId"] for node in core_nodes]
    print(f"Found core instances: {core_instance_ids}.")

    commands = [
        # Copy the shell script from Amazon S3 to each node instance.
```

```

    f"aws s3 cp {script_path} /home/hadoop",
    # Run the shell script to install libraries on each node instance.
    "bash /home/hadoop/install_libraries.sh",
]
for command in commands:
    print(f"Sending '{command}' to core instances...")
    command_id = ssm_client.send_command(
        InstanceIds=core_instance_ids,
        DocumentName="AWS-RunShellScript",
        Parameters={"commands": [command]},
        TimeoutSeconds=3600,
    )["Command"]["CommandId"]
    while True:
        # Verify the previous step succeeded before running the next step.
        cmd_result = ssm_client.list_commands(CommandId=command_id)["Commands"][0]
        if cmd_result["StatusDetails"] == "Success":
            print(f"Command succeeded.")
            break
        elif cmd_result["StatusDetails"] in ["Pending", "InProgress"]:
            print(f"Command status is {cmd_result['StatusDetails']}, waiting...")
            time.sleep(10)
        else:
            print(f"Command status is {cmd_result['StatusDetails']}, quitting.")
            raise RuntimeError(
                f"Command {command} failed to run. "
                f"Details: {cmd_result['StatusDetails']}"
            )

def main():
    parser = argparse.ArgumentParser()
    parser.add_argument("cluster_id", help="The ID of the cluster.")
    parser.add_argument("script_path", help="The path to the script in Amazon S3.")
    args = parser.parse_args()

    emr_client = boto3.client("emr")
    ssm_client = boto3.client("ssm")

    install_libraries_on_core_nodes(
        args.cluster_id, args.script_path, emr_client, ssm_client
    )

if __name__ == "__main__":

```

main()

## JupyterHub Versionsverlauf

In der folgenden Tabelle sind die Versionen von aufgeführt, die in jeder Release-Version von Amazon EMR JupyterHub enthalten sind, zusammen mit den Komponenten, die mit der Anwendung installiert wurden. Informationen zu den Komponentenversionen in den einzelnen Versionen finden Sie im Abschnitt [Komponentenversion für Ihre Version](#) in [Amazon-EMR-7.x-Versionen](#), [Amazon-EMR-6.x-Versionen](#) oder [Amazon-EMR-5.x-Versionen](#).

### JupyterHub Versionsinformationen

Amazon-EMR-Version sbezeichnung	JupyterHub Version	Komponenten wurden installiert mit JupyterHub
emr-5.36.2	1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-7.1.0	1.5.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-

Amazon-EMR-Version sbezeichnung	JupyterHub Version	Komponenten wurden installiert mit JupyterHub
		client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-7.0.0	1.5.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-6.15.0	1.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Amazon-EMR-Version sbezeichnung	JupyterHub Version	Komponenten wurden installiert mit JupyterHub
emr-6.14.0	1.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-6.13.0	1.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub



Amazon-EMR-Version sbezeichnung	JupyterHub Version	Komponenten wurden installiert mit JupyterHub
emr-6.12.0	1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-6.11.1	1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Amazon-EMR-Version sbezeichnung	JupyterHub Version	Komponenten wurden installiert mit JupyterHub
emr-6.11.0	1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-6.10.1	1.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Amazon-EMR-Version sbezeichnung	JupyterHub Version	Komponenten wurden installiert mit JupyterHub
emr-6.10.0	1.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-6.9.1	1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Amazon-EMR-Version sbezeichnung	JupyterHub Version	Komponenten wurden installiert mit JupyterHub
emr-6.9.0	1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-6.8.1	1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Amazon-EMR-Version sbezeichnung	JupyterHub Version	Komponenten wurden installiert mit JupyterHub
emr-6.8.0	1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-6.7.0	1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Amazon-EMR-Version sbezeichnung	JupyterHub Version	Komponenten wurden installiert mit JupyterHub
emr-5.36.1	1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.36.0	1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Amazon-EMR-Version sbezeichnung	JupyterHub Version	Komponenten wurden installiert mit JupyterHub
emr-6.6.0	1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.35.0	1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Amazon-EMR-Version sbezeichnung	JupyterHub Version	Komponenten wurden installiert mit JupyterHub
emr-6.5.0	1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-6.4.0	1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub



Amazon-EMR-Version sbezeichnung	JupyterHub Version	Komponenten wurden installiert mit JupyterHub
emr-6.3.1	1.2.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-6.3.0	1.2.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Amazon-EMR-Version sbezeichnung	JupyterHub Version	Komponenten wurden installiert mit JupyterHub
emr-6.2.1	1.1.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-6.2.0	1.1.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Amazon-EMR-Version sbezeichnung	JupyterHub Version	Komponenten wurden installiert mit JupyterHub
emr-6.1.1	1.1.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-6.1.0	1.1.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Amazon-EMR-Version sbezeichnung	JupyterHub Version	Komponenten wurden installiert mit JupyterHub
emr-6.0.1	1.0.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-6.0.0	1.0.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Amazon-EMR-Version sbezeichnung	JupyterHub Version	Komponenten wurden installiert mit JupyterHub
emr-5.34.0	1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.33.1	1.2.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Amazon-EMR-Version sbezeichnung	JupyterHub Version	Komponenten wurden installiert mit JupyterHub
emr-5.33.0	1.2.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.32.1	1.1.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Amazon-EMR-Version sbezeichnung	JupyterHub Version	Komponenten wurden installiert mit JupyterHub
emr-5.32.0	1.1.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.31.1	1.1.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Amazon-EMR-Version sbezeichnung	JupyterHub Version	Komponenten wurden installiert mit JupyterHub
emr-5.31.0	1.1.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.30.2	1.1.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub



Amazon-EMR-Version sbezeichnung	JupyterHub Version	Komponenten wurden installiert mit JupyterHub
emr-5.30.1	1.1.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.30.0	1.1.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Amazon-EMR-Version sbezeichnung	JupyterHub Version	Komponenten wurden installiert mit JupyterHub
emr-5.29.0	1.0.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.28.1	1.0.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Amazon-EMR-Version sbezeichnung	JupyterHub Version	Komponenten wurden installiert mit JupyterHub
emr-5.28.0	1.0.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.27.1	1.0.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Amazon-EMR-Version sbezeichnung	JupyterHub Version	Komponenten wurden installiert mit JupyterHub
emr-5.27.0	1.0.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.26.0	0.9.6	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Amazon-EMR-Version sbezeichnung	JupyterHub Version	Komponenten wurden installiert mit JupyterHub
emr-5.25.0	0.9.6	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.24.1	0.9.6	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Amazon-EMR-Version sbezeichnung	JupyterHub Version	Komponenten wurden installiert mit JupyterHub
emr-5.24.0	0.9.6	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.23.1	0.9.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Amazon-EMR-Version sbezeichnung	JupyterHub Version	Komponenten wurden installiert mit JupyterHub
emr-5.23.0	0.9.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.22.0	0.9.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Amazon-EMR-Version sbezeichnung	JupyterHub Version	Komponenten wurden installiert mit JupyterHub
emr-5.21.2	0.9.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.21.1	0.9.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub



Amazon-EMR-Version sbezeichnung	JupyterHub Version	Komponenten wurden installiert mit JupyterHub
emr-5.21.0	0.9.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.20.1	0.9.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Amazon-EMR-Version sbezeichnung	JupyterHub Version	Komponenten wurden installiert mit JupyterHub
emr-5.20.0	0.9.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.19.1	0.9.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Amazon-EMR-Version sbezeichnung	JupyterHub Version	Komponenten wurden installiert mit JupyterHub
emr-5.19.0	0.9.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.18.1	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Amazon-EMR-Version sbezeichnung	JupyterHub Version	Komponenten wurden installiert mit JupyterHub
emr-5.18.0	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.17.2	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Amazon-EMR-Version sbezeichnung	JupyterHub Version	Komponenten wurden installiert mit JupyterHub
emr-5.17.1	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.17.0	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Amazon-EMR-Version sbezeichnung	JupyterHub Version	Komponenten wurden installiert mit JupyterHub
emr-5.16.1	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.16.0	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Amazon-EMR-Version sbezeichnung	JupyterHub Version	Komponenten wurden installiert mit JupyterHub
emr-5.15.1	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.15.0	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Amazon-EMR-Version sbezeichnung	JupyterHub Version	Komponenten wurden installiert mit JupyterHub
emr-5.14.2	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.14.1	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub



Amazon-EMR-Version sbezeichnung	JupyterHub Version	Komponenten wurden installiert mit JupyterHub
emr-5.14.0	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

# Apache Livy

Livy ermöglicht die Interaktion über eine REST-Schnittstelle mit einem EMR-Cluster, auf dem Spark ausgeführt wird. Sie können die REST-Schnittstelle oder eine RPC-Client-Bibliothek verwenden, um Spark-Jobs oder Bausteine von Spark-Code zu senden, Ergebnisse synchron oder asynchron abzurufen und Spark-Kontext zu verwalten. Weitere Informationen finden Sie auf der [Apache Livy-Website](#). Livy ist in Amazon-EMR-Version 5.9.0 und höher enthalten.

Richten Sie für den Zugriff auf die Livy-Webschnittstelle einen SSH-Tunnel zum Master-Knoten und eine Proxy-Verbindung ein. Weitere Informationen finden Sie unter [Auf Amazon-EMR-Clustern gehostete Webschnittstellen anzeigen](#).

Die folgende Tabelle listet die Version von Livy auf, die in der neuesten Version der Amazon-EMR-7.x-Serie enthalten ist, zusammen mit den Komponenten, die Amazon EMR mit Livy installiert.

Informationen zur Version der Komponenten, die in dieser Version mit Livy installiert wurden, finden Sie unter [Komponentenversionen von Version 7.1.0](#).

Livy-Versionsinformationen für emr-7.1.0

Amazon-EMR-Version sbezeichnung	Livy-Version	Mit Livy installierte Komponent en
emr-7.1.0	Livy 0.8.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Die folgende Tabelle listet die Version von Livy auf, die in der neuesten Version der Amazon-EMR-6.x-Serie enthalten ist, zusammen mit den Komponenten, die Amazon EMR mit Livy installiert.

Die Version der Komponenten, die mit Livy in dieser Version installiert wurden, finden Sie unter [Komponentenversionen der Version 6.15.0](#).

Livy-Versionsinformationen für emr-6.15.0

Amazon-EMR-Version sbezeichnung	Livy-Version	Mit Livy installierte Komponent en
emr-6.15.0	Livy 0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server , hudi, hudi-spark, r, spark- client, spark-history-server, spark-on-yarn, spark-yarn- slave, livy-server, nginx

Die folgende Tabelle listet die Version von Livy auf, die in der neuesten Version der Amazon-EMR-5.x-Serie enthalten ist, zusammen mit den Komponenten, die Amazon EMR mit Livy installiert.

[Informationen zur Version der Komponenten, die in dieser Version mit Livy installiert wurden, finden Sie unter Komponentenversionen von Version 5.36.2.](#)

Livy-Versionsinformationen für emr-5.36.2

Amazon-EMR-Version sbezeichnung	Livy-Version	Mit Livy installierte Komponent en
emr-5.36.2	Livy 0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-

Amazon-EMR-Version sbezeichnung	Livy-Version	Mit Livy installierte Komponenten
		yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

## Themen

- [HTTPS mit Apache Livy aktivieren](#)
- [Livy-Versionsverlauf](#)

## HTTPS mit Apache Livy aktivieren

1. Stellen Sie einen Amazon-EMR-Cluster mit aktivierter Übertragungsverschlüsselung bereit. Weitere Informationen zur Verschlüsselung finden Sie unter [Verschlüsseln von Daten im Ruhezustand und bei der Übertragung](#).
2. Erstellen Sie eine Datei mit dem Namen `livy_ssl.sh` und den folgenden Inhalten.

```
#!/bin/bash

KEYSTORE_FILE=`awk '/ssl.server.keystore.location/{getline; print}' /etc/hadoop/conf/ssl-server.xml | sed -e 's/<[^>]*>//g' | tr -d ' \t\n\r\f'`
KEYSTORE_PASS=`awk '/ssl.server.keystore.password/{getline; print}' /etc/hadoop/conf/ssl-server.xml | sed -e 's/<[^>]*>//g' | tr -d ' \t\n\r\f'`
KEY_PASS=`awk '/ssl.server.keystore.keypassword/{getline; print}' /etc/hadoop/conf/ssl-server.xml | sed -e 's/<[^>]*>//g' | tr -d ' \t\n\r\f'`

echo "livy.keystore $KEYSTORE_FILE
livy.keystore.password $KEYSTORE_PASS
livy.key-password $KEY_PASS" | sudo tee -a /etc/livy/conf/livy.conf >/dev/null

sudo systemctl restart livy-server.service
```

3. Führen Sie das folgende Skript als Amazon-EMR-Schritt aus. Dieses Skript ändert `/etc/livy/conf/livy.conf`, um SSL zu aktivieren.

```
--steps '[{"Args":["s3://DOC-EXAMPLE-BUCKET/livy_ssl.sh"],"Type":"CUSTOM_JAR","ActionOnFailure":"CONTINUE","Jar":"s3://us-east-1.elasticmapreduce/libs/script-runner/script-runner.jar","Properties":"","Name":"Custom JAR"}]'
```

4. Starten Sie den Apache-Livy-Service neu, damit die Änderungen wirksam werden. Informationen zum Neustarten von Apache Livy finden Sie unter [Stoppen und Neustarten von Prozessen](#).
5. Testen Sie, ob die Clients jetzt über HTTPS kommunizieren können. Um beispielsweise einen Job einzureichen, führen Sie den folgenden Code aus.

```
curl -k -X POST --data '{"file": "local:///usr/lib/spark/examples/jars/spark-examples.jar",
"className": "org.apache.spark.examples.SparkPi"}' \
-H "Content-Type: application/json" \
https://EMR_Master_Node_Host:8998/batches
```

Wenn Sie HTTPS erfolgreich aktiviert haben, sendet Livy eine Antwort, in der angegeben wird, dass der Befehl akzeptiert und der Batch-Job gesendet wurde.

```
{"id":1,"name":null,"owner":null,"proxyUser":null,"state":"starting","appId":null,"appInfo":{"driverLogUrl":null,"sparkUiUrl":null},"log":["stdout: ", "\nstderr: ", "\nYARN Diagnostics: "]}
```

## Livy-Versionsverlauf

In der folgenden Tabelle sind die Version von Livy aufgeführt, die in jeder Release-Version von Amazon EMR enthalten ist, zusammen mit den Komponenten, die mit der Anwendung installiert wurden. Informationen zu den Komponentenversionen in den einzelnen Versionen finden Sie im Abschnitt [Komponentenversion für Ihre Version in Amazon-EMR-7.x-Versionen](#), [Amazon-EMR-6.x-Versionen](#) oder [Amazon-EMR-5.x-Versionen](#).

## Die Livy Versionsinformationen

Amazon-EMR-Version sbezeichnung	Livy-Version	Mit Livy installierte Komponent en
emr-5.36,2	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server , hudi, hudi-spark, r, spark- client, spark-history-server, spark-on-yarn, spark-yarn- slave, livy-server, nginx
emr-7.1.0	0.8.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server , hudi, hudi-spark, r, spark- client, spark-history-server, spark-on-yarn, spark-yarn- slave, livy-server, nginx
emr-7.0.0	0.7.1	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop-

Amazon-EMR-Version sbezeichnung	Livy-Version	Mit Livy installierte Komponent en
		yarn-resourcemanager, hadoop-yarn-timeline-server , hudi, hudi-spark, r, spark- client, spark-history-server, spark-on-yarn, spark-yarn- slave, livy-server, nginx
emr-6.15.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server , hudi, hudi-spark, r, spark- client, spark-history-server, spark-on-yarn, spark-yarn- slave, livy-server, nginx
emr-6.14.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server , hudi, hudi-spark, r, spark- client, spark-history-server, spark-on-yarn, spark-yarn- slave, livy-server, nginx

Amazon-EMR-Version sbezeichnung	Livy-Version	Mit Livy installierte Komponent en
emr-6.13.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server , hudi, hudi-spark, r, spark- client, spark-history-server, spark-on-yarn, spark-yarn- slave, livy-server, nginx
emr-6.12.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server , hudi, hudi-spark, r, spark- client, spark-history-server, spark-on-yarn, spark-yarn- slave, livy-server, nginx



Amazon-EMR-Version sbezeichnung	Livy-Version	Mit Livy installierte Komponent en
emr-6.11.1	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server , hudi, hudi-spark, r, spark- client, spark-history-server, spark-on-yarn, spark-yarn- slave, livy-server, nginx
emr-6.11.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server , hudi, hudi-spark, r, spark- client, spark-history-server, spark-on-yarn, spark-yarn- slave, livy-server, nginx

Amazon-EMR-Version sbezeichnung	Livy-Version	Mit Livy installierte Komponent en
emr-6.10.1	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server , hudi, hudi-spark, r, spark- client, spark-history-server, spark-on-yarn, spark-yarn- slave, livy-server, nginx
emr-6.10.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server , hudi, hudi-spark, r, spark- client, spark-history-server, spark-on-yarn, spark-yarn- slave, livy-server, nginx

Amazon-EMR-Version sbezeichnung	Livy-Version	Mit Livy installierte Komponent en
emr-6.9.1	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-6.9.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Amazon-EMR-Version sbezeichnung	Livy-Version	Mit Livy installierte Komponent en
emr-6.8.1	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server , hudi, hudi-spark, r, spark- client, spark-history-server, spark-on-yarn, spark-yarn- slave, livy-server, nginx
emr-6.8.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server , hudi, hudi-spark, r, spark- client, spark-history-server, spark-on-yarn, spark-yarn- slave, livy-server, nginx

Amazon-EMR-Version sbezeichnung	Livy-Version	Mit Livy installierte Komponent en
emr-6.7.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.36.1	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Amazon-EMR-Version sbezeichnung	Livy-Version	Mit Livy installierte Komponent en
emr-5.36.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server , hudi, hudi-spark, r, spark- client, spark-history-server, spark-on-yarn, spark-yarn- slave, livy-server, nginx
emr-6.6.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server , hudi, hudi-spark, r, spark- client, spark-history-server, spark-on-yarn, spark-yarn- slave, livy-server, nginx

Amazon-EMR-Version sbezeichnung	Livy-Version	Mit Livy installierte Komponent en
emr-5.35.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-serv er, spark-on-yarn, spark-yarn- slave, livy-server, nginx
emr-6.5.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-serv er, spark-on-yarn, spark-yarn- slave, livy-server, nginx

Amazon-EMR-Version sbezeichnung	Livy-Version	Mit Livy installierte Komponent en
emr-6.4.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-serv er, spark-on-yarn, spark-yarn- slave, livy-server, nginx
emr-6.3.1	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-serv er, spark-on-yarn, spark-yarn- slave, livy-server, nginx



Amazon-EMR-Version sbezeichnung	Livy-Version	Mit Livy installierte Komponent en
emr-6.3.0	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-serv er, spark-on-yarn, spark-yarn- slave, livy-server, nginx
emr-6.2.1	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-serv er, spark-on-yarn, spark-yarn- slave, livy-server, nginx

Amazon-EMR-Version sbezeichnung	Livy-Version	Mit Livy installierte Komponent en
emr-6.2.0	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-serv er, spark-on-yarn, spark-yarn- slave, livy-server, nginx
emr-6.1.1	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-serv er, spark-on-yarn, spark-yarn- slave, livy-server, nginx

Amazon-EMR-Version sbezeichnung	Livy-Version	Mit Livy installierte Komponent en
emr-6.1.0	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-serv er, spark-on-yarn, spark-yarn- slave, livy-server, nginx
emr-6.0.1	0.6.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-serv er, spark-on-yarn, spark-yarn- slave, livy-server, nginx

Amazon-EMR-Version sbezeichnung	Livy-Version	Mit Livy installierte Komponent en
emr-6.0.0	0.6.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-serv er, spark-on-yarn, spark-yarn- slave, livy-server, nginx
emr-5.34.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-serv er, spark-on-yarn, spark-yarn- slave, livy-server, nginx

Amazon-EMR-Version sbezeichnung	Livy-Version	Mit Livy installierte Komponent en
emr-5.33.1	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-serv er, spark-on-yarn, spark-yarn- slave, livy-server, nginx
emr-5.33.0	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-serv er, spark-on-yarn, spark-yarn- slave, livy-server, nginx

Amazon-EMR-Version sbezeichnung	Livy-Version	Mit Livy installierte Komponent en
emr-5.32.1	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-serv er, spark-on-yarn, spark-yarn- slave, livy-server, nginx
emr-5.32.0	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-serv er, spark-on-yarn, spark-yarn- slave, livy-server, nginx

Amazon-EMR-Version sbezeichnung	Livy-Version	Mit Livy installierte Komponent en
emr-5.31.1	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-serv er, spark-on-yarn, spark-yarn- slave, livy-server, nginx
emr-5.31.0	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-serv er, spark-on-yarn, spark-yarn- slave, livy-server, nginx

Amazon-EMR-Version sbezeichnung	Livy-Version	Mit Livy installierte Komponent en
emr-5.30.2	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-serv er, spark-on-yarn, spark-yarn- slave, livy-server, nginx
emr-5.30.1	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-serv er, spark-on-yarn, spark-yarn- slave, livy-server, nginx



Amazon-EMR-Version sbezeichnung	Livy-Version	Mit Livy installierte Komponent en
emr-5.30.0	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-notebook-env, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.29.0	0.6.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Amazon-EMR-Version sbezeichnung	Livy-Version	Mit Livy installierte Komponent en
emr-5.28.1	0.6.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-serv er, spark-on-yarn, spark-yarn- slave, livy-server, nginx
emr-5.28.0	0.6.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-serv er, spark-on-yarn, spark-yarn- slave, livy-server, nginx

Amazon-EMR-Version sbezeichnung	Livy-Version	Mit Livy installierte Komponent en
emr-5.27.1	0.6.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-serv er, spark-on-yarn, spark-yarn- slave, livy-server, nginx
emr-5.27.0	0.6.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-serv er, spark-on-yarn, spark-yarn- slave, livy-server, nginx

Amazon-EMR-Version sbezeichnung	Livy-Version	Mit Livy installierte Komponent en
emr-5.26.0	0.6.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-serv er, spark-on-yarn, spark-yarn- slave, livy-server, nginx
emr-5.25.0	0.6.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-serv er, spark-on-yarn, spark-yarn- slave, livy-server, nginx

Amazon-EMR-Version sbezeichnung	Livy-Version	Mit Livy installierte Komponent en
emr-5.24.1	0.6.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-serv er, spark-on-yarn, spark-yarn- slave, livy-server, nginx
emr-5.24.0	0.6.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-serv er, spark-on-yarn, spark-yarn- slave, livy-server, nginx

Amazon-EMR-Version sbezeichnung	Livy-Version	Mit Livy installierte Komponent en
emr-5.23.1	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-serv er, spark-on-yarn, spark-yarn- slave, livy-server, nginx
emr-5.23.0	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-serv er, spark-on-yarn, spark-yarn- slave, livy-server, nginx

Amazon-EMR-Version sbezeichnung	Livy-Version	Mit Livy installierte Komponent en
emr-5.22.0	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-serv er, spark-on-yarn, spark-yarn- slave, livy-server, nginx
emr-5.21.2	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-serv er, spark-on-yarn, spark-yarn- slave, livy-server, nginx

Amazon-EMR-Version sbezeichnung	Livy-Version	Mit Livy installierte Komponent en
emr-5.21.1	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-serv er, spark-on-yarn, spark-yarn- slave, livy-server, nginx
emr-5.21.0	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-serv er, spark-on-yarn, spark-yarn- slave, livy-server, nginx



Amazon-EMR-Version sbezeichnung	Livy-Version	Mit Livy installierte Komponent en
emr-5.20.1	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-serv er, spark-on-yarn, spark-yarn- slave, livy-server, nginx
emr-5.20.0	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-serv er, spark-on-yarn, spark-yarn- slave, livy-server, nginx

Amazon-EMR-Version sbezeichnung	Livy-Version	Mit Livy installierte Komponent en
emr-5.19.1	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-serv er, spark-on-yarn, spark-yarn- slave, livy-server, nginx
emr-5.19.0	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-serv er, spark-on-yarn, spark-yarn- slave, livy-server, nginx

Amazon-EMR-Version sbezeichnung	Livy-Version	Mit Livy installierte Komponent en
emr-5.18.1	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-serv er, spark-on-yarn, spark-yarn- slave, livy-server, nginx
emr-5.18.0	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-serv er, spark-on-yarn, spark-yarn- slave, livy-server, nginx

Amazon-EMR-Version sbezeichnung	Livy-Version	Mit Livy installierte Komponent en
emr-5.17.2	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-serv er, spark-on-yarn, spark-yarn- slave, livy-server
emr-5.17.1	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-serv er, spark-on-yarn, spark-yarn- slave, livy-server

Amazon-EMR-Version sbezeichnung	Livy-Version	Mit Livy installierte Komponent en
emr-5.17.0	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-serv er, spark-on-yarn, spark-yarn- slave, livy-server
emr-5.16.1	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-serv er, spark-on-yarn, spark-yarn- slave, livy-server

Amazon-EMR-Version sbezeichnung	Livy-Version	Mit Livy installierte Komponent en
emr-5.16.0	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-serv er, spark-on-yarn, spark-yarn- slave, livy-server
emr-5.15.1	0,4,0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-serv er, spark-on-yarn, spark-yarn- slave, livy-server

Amazon-EMR-Version sbezeichnung	Livy-Version	Mit Livy installierte Komponent en
emr-5.15.0	0,4,0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-serv er, spark-on-yarn, spark-yarn- slave, livy-server
emr-5.14.2	0,4,0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-serv er, spark-on-yarn, spark-yarn- slave, livy-server

Amazon-EMR-Version sbezeichnung	Livy-Version	Mit Livy installierte Komponent en
emr-5.14.1	0,4,0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-serv er, spark-on-yarn, spark-yarn- slave, livy-server
emr-5.14.0	0,4,0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-serv er, spark-on-yarn, spark-yarn- slave, livy-server



Amazon-EMR-Version sbezeichnung	Livy-Version	Mit Livy installierte Komponent en
emr-5.13.1	0,4,0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-serv er, spark-on-yarn, spark-yarn- slave, livy-server
emr-5.13.0	0,4,0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-serv er, spark-on-yarn, spark-yarn- slave, livy-server

Amazon-EMR-Version sbezeichnung	Livy-Version	Mit Livy installierte Komponent en
emr-5.12.3	0,4,0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-serv er, spark-on-yarn, spark-yarn- slave, livy-server
emr-5.12.2	0,4,0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-serv er, spark-on-yarn, spark-yarn- slave, livy-server

Amazon-EMR-Version sbezeichnung	Livy-Version	Mit Livy installierte Komponent en
emr-5.12.1	0,4,0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-serv er, spark-on-yarn, spark-yarn- slave, livy-server
emr-5.12.0	0,4,0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-serv er, spark-on-yarn, spark-yarn- slave, livy-server

Amazon-EMR-Version sbezeichnung	Livy-Version	Mit Livy installierte Komponent en
emr-5.11.4	0,4,0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-serv er, spark-on-yarn, spark-yarn- slave, livy-server
emr-5.11.3	0,4,0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-serv er, spark-on-yarn, spark-yarn- slave, livy-server

Amazon-EMR-Version sbezeichnung	Livy-Version	Mit Livy installierte Komponent en
emr-5.11.2	0,4,0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-serv er, spark-on-yarn, spark-yarn- slave, livy-server
emr-5.11.1	0,4,0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-serv er, spark-on-yarn, spark-yarn- slave, livy-server

Amazon-EMR-Version sbezeichnung	Livy-Version	Mit Livy installierte Komponent en
emr-5.11.0	0,4,0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-serv er, spark-on-yarn, spark-yarn- slave, livy-server
emr-5.10.1	0,4,0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-serv er, spark-on-yarn, spark-yarn- slave, livy-server

Amazon-EMR-Version sbezeichnung	Livy-Version	Mit Livy installierte Komponent en
emr-5.10.0	0,4,0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server
emr-5.9.1	0,4,0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server

Amazon-EMR-Version sbezeichnung	Livy-Version	Mit Livy installierte Komponent en
emr-5.9.0	0,4,0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs- datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-kms-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-serv er, spark-on-yarn, spark-yarn- slave, livy-server



# Apache MXNet

Apache MXNet ist eine Beschleunigungsbibliothek, die auf die Erstellung neuronaler Netzwerke und anderer Deep Learning-Anwendungen ausgelegt ist. MXNet automatisiert allgemeine Arbeitsabläufe und optimiert numerische Berechnungen. Mit MXNet können Sie neuronale Netzwerkarchitekturen entwerfen, ohne dass Sie sich auf die Implementierung von einfachen Berechnungen konzentrieren müssen, wie z. B. lineare Algebra-Operationen. MXNet ist in Amazon-EMR-Version 5.10.0 und höher enthalten.

Weitere Informationen finden Sie auf der [Apache MXNet-Website](#).

Die folgende Tabelle listet die Version von MXNet auf, die in der neuesten Version der Amazon-EMR-7.x-Serie enthalten ist, zusammen mit den Komponenten, die Amazon EMR mit MXNet installiert.

Die Version der Komponenten, die in dieser Version mit MXNet installiert wurden, finden Sie unter [Komponentenversionen von Version 7.1.0](#).

MXNet-Versionsinformationen für emr-7.1.0

Amazon-EMR-Version sbezeichnung	MXNet-Version	Mit MXNet installierte Komponenten
emr-7.1.0	MXNet 1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Die folgende Tabelle listet die Version von MXNet auf, die in der neuesten Version der Amazon-EMR-6.x-Serie enthalten ist, zusammen mit den Komponenten, die Amazon EMR mit MXNet installiert.

Die Version der Komponenten, die mit MXNet in dieser Version installiert wurden, finden Sie unter [Komponentenversionen der Version 6.15.0](#).

#### MXNet-Versionsinformationen für emr-6.15.0

Amazon-EMR-Version sbezeichnung	MXNet-Version	Mit MXNet installierte Komponenten
emr-6.15.0	MXNet 1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Die folgende Tabelle listet die Version von MXNet auf, die in der neuesten Version der Amazon-EMR-5.x-Serie enthalten ist, zusammen mit den Komponenten, die Amazon EMR mit MXNet installiert.

Die Version der Komponenten, die in dieser Version mit MXNet installiert wurden, finden Sie unter [Version 5.36.2](#) Komponentenversionen.

#### MXNet-Versionsinformationen für emr-5.36.2

Amazon-EMR-Version sbezeichnung	MXNet-Version	Mit MXNet installierte Komponenten
emr-5.36.2	MXNet 1.8.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resour

Amazon-EMR-Version sbezeichnung	MXNet-Version	Mit MXNet installierte Komponenten
		cemanager, hadoop-yarn- timeline-server, mxnet, opencv

## MXNet-Versionsverlauf

In der folgenden Tabelle sind die Version von MXNet aufgeführt, die in jeder Release-Version von Amazon EMR enthalten ist, zusammen mit den Komponenten, die mit der Anwendung installiert wurden. Informationen zu den Komponentenversionen in den einzelnen Versionen finden Sie im Abschnitt [Komponentenversion für Ihre Version in Amazon-EMR-7.x-Versionen](#), [Amazon-EMR-6.x-Versionen](#) oder [Amazon-EMR-5.x-Versionen](#).

Die MXNet Versionsinformationen

Amazon-EMR-Version sbezeichnung	MXNet-Version	Mit MXNet installierte Komponenten
emr-5.36,2	1.8.0	emrfs, emr-goodies, hadoop- client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, mxnet, opencv
emr-7.1.0	1.9.1	emrfs, emr-goodies, hadoop- client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour

Amazon-EMR-Version sbezeichnung	MXNet-Version	Mit MXNet installierte Komponenten
		cemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-7.0.0	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.15.0	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.14.0	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Amazon-EMR-Version sbezeichnung	MXNet-Version	Mit MXNet installierte Komponenten
emr-6.13.0	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.12.0	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.11.1	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv

Amazon-EMR-Version sbezeichnung	MXNet-Version	Mit MXNet installierte Komponenten
emr-6.11.0	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.10.1	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.10.0	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Amazon-EMR-Version sbezeichnung	MXNet-Version	Mit MXNet installierte Komponenten
emr-6.9.1	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.9.0	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.8.1	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv

Amazon-EMR-Version sbezeichnung	MXNet-Version	Mit MXNet installierte Komponenten
emr-6.8.0	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.7.0	1.8.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.36.1	1.8.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv



Amazon-EMR-Version sbezeichnung	MXNet-Version	Mit MXNet installierte Komponenten
emr-5.36.0	1.8.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.6.0	1.8.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.35.0	1.8.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Amazon-EMR-Version sbezeichnung	MXNet-Version	Mit MXNet installierte Komponenten
emr-6.5.0	1.8.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.4.0	1.8.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.3.1	1.7.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Amazon-EMR-Version sbezeichnung	MXNet-Version	Mit MXNet installierte Komponenten
emr-6.3.0	1.7.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.2.1	1.7.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.2.0	1.7.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Amazon-EMR-Version sbezeichnung	MXNet-Version	Mit MXNet installierte Komponenten
emr-6.1.1	1.6.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.1.0	1.6.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.0.1	1.5.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Amazon-EMR-Version sbezeichnung	MXNet-Version	Mit MXNet installierte Komponenten
emr-6.0.0	1.5.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.34.0	1.8.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.33.1	1.7.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Amazon-EMR-Version sbezeichnung	MXNet-Version	Mit MXNet installierte Komponenten
emr-5.33.0	1.7.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.32.1	1.7.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.32.0	1.7.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Amazon-EMR-Version sbezeichnung	MXNet-Version	Mit MXNet installierte Komponenten
emr-5.31.1	1.6.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.31.0	1.6.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.30.2	1.5.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Amazon-EMR-Version sbezeichnung	MXNet-Version	Mit MXNet installierte Komponenten
emr-5.30.1	1.5.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.30.0	1.5.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.29.0	1.5.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv



Amazon-EMR-Version sbezeichnung	MXNet-Version	Mit MXNet installierte Komponenten
emr-5.28.1	1.5.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.28.0	1.5.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.27.1	1.4.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Amazon-EMR-Version sbezeichnung	MXNet-Version	Mit MXNet installierte Komponenten
emr-5.27.0	1.4.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.26.0	1.4.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.25.0	1.4.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Amazon-EMR-Version sbezeichnung	MXNet-Version	Mit MXNet installierte Komponenten
emr-5.24.1	1.4.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.24.0	1.4.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.23.1	1.3.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv

Amazon-EMR-Version sbezeichnung	MXNet-Version	Mit MXNet installierte Komponenten
emr-5.23.0	1.3.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.22.0	1.3.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.21.2	1.3.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Amazon-EMR-Version sbezeichnung	MXNet-Version	Mit MXNet installierte Komponenten
emr-5.21.1	1.3.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.21.0	1.3.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.20.1	1.3.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv

Amazon-EMR-Version sbezeichnung	MXNet-Version	Mit MXNet installierte Komponenten
emr-5.20.0	1.3.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.19.1	1.3.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.19.0	1.3.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv

Amazon-EMR-Version sbezeichnung	MXNet-Version	Mit MXNet installierte Komponenten
emr-5.18.1	1.2.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.18.0	1.2.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.17.2	1.2.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Amazon-EMR-Version sbezeichnung	MXNet-Version	Mit MXNet installierte Komponenten
emr-5.17.1	1.2.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.17.0	1.2.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.16.1	1.2.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv



Amazon-EMR-Version sbezeichnung	MXNet-Version	Mit MXNet installierte Komponenten
emr-5.16.0	1.2.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.15.1	1.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.15.0	1.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Amazon-EMR-Version sbezeichnung	MXNet-Version	Mit MXNet installierte Komponenten
emr-5.14.2	1.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.14.1	1.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.14.0	1.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Amazon-EMR-Version sbezeichnung	MXNet-Version	Mit MXNet installierte Komponenten
emr-5.13.1	1.0.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet
emr-5.13.0	1.0.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet
emr-5.12.3	1.0.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet

Amazon-EMR-Version sbezeichnung	MXNet-Version	Mit MXNet installierte Komponenten
emr-5.12.2	1.0.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet
emr-5.12.1	1.0.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet
emr-5.12.0	1.0.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet

Amazon-EMR-Version sbezeichnung	MXNet-Version	Mit MXNet installierte Komponenten
emr-5.11.4	0.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet
emr-5.11.3	0.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet
emr-5.11.2	0.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet

Amazon-EMR-Version sbezeichnung	MXNet-Version	Mit MXNet installierte Komponenten
emr-5.11.1	0.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet
emr-5.11.0	0.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet
emr-5.10.1	0.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet

Amazon-EMR-Version sbezeichnung	MXNet-Version	Mit MXNet installierte Komponenten
emr-5.10.0	0.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet

# Apache Oozie

Verwenden Sie den Apache Oozie Workflow Scheduler, um Hadoop-Aufträge zu verwalten und zu koordinieren. Weitere Informationen finden Sie unter <http://oozie.apache.org/>.

Die native Weboberfläche von Oozie wird in Amazon EMR nicht unterstützt. Wenn Sie eine Front-End-Oberfläche für Oozie nutzen möchten, können Sie die Hue Oozie-Anwendung verwenden. Weitere Informationen finden Sie unter [Hue](#). Oozie ist in Amazon-EMR-Version 5.0.0 und höher enthalten. Oozie ist als Sandbox-Anwendung in früheren Versionen enthalten. Weitere Informationen finden Sie unter [Amazon-EMR-4.x-Versionen](#).

Wenn Sie ein benutzerdefiniertes Amazon-Linux-AMI verwenden, das auf einem Amazon-Linux-AMI mit einem Erstellungsdatum vom 8.11.2018 basiert, kann der Oozie-Server nicht gestartet werden. Wenn Sie Oozie verwenden, erstellen Sie ein benutzerdefiniertes AMI, das auf einer Amazon-Linux-AMI-ID mit einem anderen Erstellungsdatum basiert. Sie können den folgenden AWS CLI Befehl verwenden, um eine Liste der Image-IDs für alle HVM-Amazon-Linux-AMIs mit einer Version 2018.03 zusammen mit dem Veröffentlichungsdatum zurückzugeben, sodass Sie ein geeignetes Amazon Linux-AMI als Basis auswählen können. MyRegion Ersetzen Sie es durch Ihre Regionskennung, z. B. us-west-2.

```
aws ec2 --region MyRegion describe-images --owner amazon --query 'Images[?
Name!=`null`][[?starts_with(Name, `amzn-ami-hvm-2018.03`) == `true`].
[CreationDate,ImageId,Name]'
```

Die folgende Tabelle listet die Version von Oozie auf, die in der neuesten Version der Amazon-EMR-7.x-Serie enthalten ist, zusammen mit den Komponenten, die Amazon EMR mit Oozie installiert.

Informationen zur Version der Komponenten, die in dieser Version mit Oozie installiert wurden, finden Sie unter [Komponentenversionen von Version 7.1.0](#).

Oozie-Versionsinformationen für emr-7.1.0

Amazon-EMR-Version sbezeichnung	Oozie-Version	Mit Oozie installierte Komponenten
emr-7.1.0	Oozie 5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-



Amazon-EMR-Version sbezeichnung	Oozie-Version	Mit Oozie installierte Komponenten
		cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn, tez-on-worker

Die folgende Tabelle listet die Version von Oozie auf, die in der neuesten Version der Amazon-EMR-6.x-Serie enthalten ist, zusammen mit den Komponenten, die Amazon EMR mit Oozie installiert.

Die Version der Komponenten, die mit Oozie in dieser Version installiert wurden, finden Sie unter [Komponentenversionen der Version 6.15.0](#).

#### Oozie-Versionsinformationen für emr-6.15.0

Amazon-EMR-Version sbezeichnung	Oozie-Version	Mit Oozie installierte Komponenten
emr-6.15.0	Oozie 5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client,

Amazon-EMR-Version sbezeichnung	Oozie-Version	Mit Oozie installierte Komponenten
		oozie-server, tez-on-yarn, tez-on-worker

Die folgende Tabelle listet die Version von Oozie auf, die in der neuesten Version der Amazon-EMR-5.x-Serie enthalten ist, zusammen mit den Komponenten, die Amazon EMR mit Oozie installiert.

[Informationen zur Version der Komponenten, die in dieser Version mit Oozie installiert wurden, finden Sie unter Komponentenversionen von Version 5.36.2.](#)

Oozie-Versionsinformationen für emr-5.36.2

Amazon-EMR-Version sbezeichnung	Oozie-Version	Mit Oozie installierte Komponenten
emr-5.36.2	Oozie 5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

## Themen

- [Oozie mit einer Remote-Datenbank in Amazon RDS verwenden](#)
- [Die Java-Version für Oozie konfigurieren](#)
- [Versionsverlauf von Oozie](#)

# Oozie mit einer Remote-Datenbank in Amazon RDS verwenden

Standardmäßig werden Benutzerinformationen und Abfrageverläufe von Oozie in einer lokalen MySQL-Datenbank auf dem Master-Knoten gespeichert. Sie können alternativ einen oder mehrere Oozie-fähige Cluster mithilfe einer in Amazon S3 gespeicherten Konfiguration und einer MySQL-Datenbank in Amazon Relational Database Service (Amazon RDS) erstellen. So können Sie Benutzerdaten und Abfrageverläufe, die von Oozie erstellt wurden, dauerhaft speichern, ohne dass der Amazon-EMR-Cluster weiter ausgeführt wird. Wir empfehlen die Verwendung der serverseitigen Amazon-S3-Verschlüsselung zum Speichern der Konfigurationsdatei.

Erstellen Sie zunächst die Remote-Datenbank für Oozie.

So erstellen Sie die externe MySQL-Datenbank

1. Öffnen Sie die Amazon-RDS-Konsole unter <https://console.aws.amazon.com/rds/>.
2. Klicken Sie auf Launch a DB Instance (DB-Instance starten).
3. Wählen Sie „MySQL“ und danach Select (Auswählen) aus.
4. Übernehmen Sie die Standardauswahl von Multi-AZ Deployment and Provisioned IOPS Storage (Multi-AZ-Bereitstellung und bereitgestellte IOPS-Speicherung) und wählen Sie Next (Weiter) aus.
5. Übernehmen Sie die Standardwerte für die Instance-Spezifikationen, geben Sie Einstellungen an und wählen Sie Next (Weiter).
6. Wählen Sie auf der Seite „Configure Advanced Settings (Erweiterte Einstellungen konfigurieren)“ geeignete Sicherheitsgruppen- und Datenbanknamen aus. Die Sicherheitsgruppe, die Sie verwenden, muss mindestens eingehenden TCP-Zugriff für Port 3306 über den Master-Knoten Ihres Clusters zulassen. Wenn Sie zu diesem Zeitpunkt noch keinen Cluster erstellt haben, können Sie für alle Hosts eine Verbindung mit Port 3306 zulassen und die Sicherheitsgruppe anpassen, nachdem Sie den Cluster gestartet haben. Klicken Sie auf DB-Instance starten.
7. Wählen Sie im RDS-Dashboard Instances und die Instance aus, die Sie soeben erstellt haben. Wenn Ihre Datenbank verfügbar ist, notieren Sie dbname, Benutzername, Kennwort und RDS-Instance. Diese Informationen benötigen Sie zum Erstellen und Konfigurieren Ihres Clusters.

Um eine externe MySQL-Datenbank für Oozie anzugeben, wenn Sie einen Cluster mit dem starten AWS CLI

Um eine externe MySQL-Datenbank für Oozie beim Start eines Clusters mithilfe der AWS CLI anzugeben, verwenden Sie die Informationen, die Sie beim Erstellen Ihrer RDS-Instance für die Konfiguration von `oozie-site` mit einem Konfigurationsobjekt notiert haben.

**Note**

Sie können mehrere Cluster mit derselben externen Datenbank erstellen, aber jeder Cluster gibt Abfrageverläufe und Benutzerdaten frei.

- Erstellen Sie mit dem einen Cluster AWS CLI, auf dem Oozie installiert ist, verwenden Sie die externe Datenbank, die Sie erstellt haben, und verweisen Sie auf eine Konfigurationsdatei mit einer Konfigurationsklassifizierung für Oozie, die die Datenbankeigenschaften spezifiziert. Im folgenden Beispiel wird ein Cluster mit installiertem Oozie erstellt, der auf die Konfigurationsdatei `myConfig.json` in Amazon S3 verweist, die die Datenbankkonfiguration angibt.

**Note**

Linux-Zeilenfortsetzungszeichen (`\`) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (`^`).

```
aws emr create-cluster --release-label emr-7.1.0 --applications Name=Oozie
Name=Spark Name=Hive \
--instance-type m5.xlarge --instance-count 3 \
--configurations https://s3.amazonaws.com/mybucket/myfolder/myConfig.json --use-
default-roles
```

Beispielinhalte der `myConfig.json`-Datei werden unten angezeigt. Ersetzen Sie *JDBC-URL*, *username* und *password* durch die JDBC-URL, den Benutzernamen und das Passwort Ihrer RDS-Instance.

**⚠ Important**

Die JDBC-URL muss den Datenbanknamen als Suffix enthalten. Zum Beispiel `jdbc:mysql://.xxxxxxxxx.us-east-1.rds.amazonaws.com:3306/ dbname. oozie-external-db`

```
[{
  "Classification": "oozie-site",
  "Properties": {
    "oozie.service.JPIService.jdbc.driver": "org.mariadb.jdbc.Driver",
    "oozie.service.JPIService.jdbc.url": "JDBC URL",

    "oozie.service.JPIService.jdbc.username": "username",
    "oozie.service.JPIService.jdbc.password": "password"
  },
  "Configurations": []
}]
```

## Die Java-Version für Oozie konfigurieren

Oozie führt mehrere Java Virtual Machine (JVM)-Prozesse aus. Auf dieser Seite wird erklärt, wie die Java-Version für jeden Prozess konfiguriert wird.

- **Oozie-Server:** Legen Sie mit `JAVA_HOME` in der `oozie-env`-Klassifizierung fest, dass die Java-Version für `EmbeddedOozieServer` aktualisiert werden soll.
- **Oozie Launcher AM:** Oozie Launcher AM ist ein Single-Mapper-MR-Job, der die entsprechenden Anwendungsbibliotheken wie Hadoop und Hive aufruft. Sofern nicht anders konfiguriert, entsprechen die Laufzeitversionen für Oozie Launcher AM den Java-Laufzeiten für Hadoop im EMR-Cluster. Um die Java-Laufzeit für Oozie Launcher AM zu konfigurieren, legen Sie für den Job die folgende `workflow.xml` Eigenschaft fest:

```
<property>
  <name>mapred.child.env</name>
  <value>JAVA_HOME=/path/to/JAVA_HOME</value>
</property>
```

Diese Eigenschaft stellt sicher, dass Oozie Launcher AM für den Oozie-Auftrag auf der von Ihnen angegebenen Java-Version ausgeführt wird und nicht auf der Java-Version, die in Hadoop festgelegt ist.

- Ausführbare Datei des Anwendungsclients: Da Oozie Launcher AM den Anwendungsclient standardmäßig aufruft, ist die Java-Laufzeit für die ausführbare Client-Datei dieselbe wie für Oozie Launcher AM.
- Von einem Oozie-Auftrag gestartete Anwendungen: Sofern nicht anders angegeben, entsprechen die Laufzeitversionen für die eigentlichen Anwendungs-JVMs, die durch einen Oozie-Auftrag gestartet werden, den Java-Laufzeiten für Hadoop im EMR-Cluster. Abhängig von der Art der Oozie-Workflow-Aktion, die verwendet wird, um die Anwendung in einem Oozie-Auftrag zu starten (Spark- oder Hive-Aktion), können Sie die Standard-Java-Laufzeit für die eigentlichen Anwendungs-JVMs im `workflow.xml` für den Oozie-Auftrag aktualisieren.

## Versionsverlauf von Oozie

In der folgenden Tabelle sind die Version von Oozie aufgeführt, die in jeder Release-Version von Amazon EMR enthalten ist, zusammen mit den Komponenten, die mit der Anwendung installiert wurden. Informationen zu den Komponentenversionen in den einzelnen Versionen finden Sie im Abschnitt [Komponentenversion für Ihre Version](#) in [Amazon-EMR-7.x-Versionen](#), [Amazon-EMR-6.x-Versionen](#) oder [Amazon-EMR-5.x-Versionen](#).

### Oozie-Versionsinformationen

Amazon-EMR-Version sbezeichnung	Oozie-Version	Mit Oozie installierte Komponenten
emr-5.36.2	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-

Amazon-EMR-Version sbezeichnung	Oozie-Version	Mit Oozie installierte Komponenten
		timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-7.1.0	5.2.1	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, oozie-client, oozie-server, tez-on-yarn, tez- on-worker
emr-7.0.0	5.2.1	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, oozie-client, oozie-server, tez-on-yarn, tez- on-worker

Amazon-EMR-Version sbezeichnung	Oozie-Version	Mit Oozie installierte Komponenten
emr-6.15.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn, tez-on-worker
emr-6.14.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn, tez-on-worker



Amazon-EMR-Version sbezeichnung	Oozie-Version	Mit Oozie installierte Komponenten
emr-6.13.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn, tez-on-worker
emr-6.12.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn, tez-on-worker

Amazon-EMR-Version sbezeichnung	Oozie-Version	Mit Oozie installierte Komponenten
emr-6.11.1	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn, tez-on-worker
emr-6.11.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn, tez-on-worker

Amazon-EMR-Version sbezeichnung	Oozie-Version	Mit Oozie installierte Komponenten
emr-6.10.1	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn, tez-on-worker
emr-6.10.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn, tez-on-worker

Amazon-EMR-Version sbezeichnung	Oozie-Version	Mit Oozie installierte Komponenten
emr-6.9.1	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-6.9.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Oozie-Version	Mit Oozie installierte Komponenten
emr-6.8.1	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-6.8.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Oozie-Version	Mit Oozie installierte Komponenten
emr-6.7.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.36.1	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Oozie-Version	Mit Oozie installierte Komponenten
emr-5.36.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-6.6.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Oozie-Version	Mit Oozie installierte Komponenten
emr-5.35.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-6.5.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn



Amazon-EMR-Version sbezeichnung	Oozie-Version	Mit Oozie installierte Komponenten
emr-6.4.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-6.3.1	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Oozie-Version	Mit Oozie installierte Komponenten
emr-6.3.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-6.2.1	5.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Oozie-Version	Mit Oozie installierte Komponenten
emr-6.2.0	5.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-6.1.1	5.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Oozie-Version	Mit Oozie installierte Komponenten
emr-6.1.0	5.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-6.0.1	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Oozie-Version	Mit Oozie installierte Komponenten
emr-6.0.0	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.34.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Oozie-Version	Mit Oozie installierte Komponenten
emr-5.33.1	5.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.33.0	5.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Oozie-Version	Mit Oozie installierte Komponenten
emr-5.32.1	5.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.32.0	5.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Oozie-Version	Mit Oozie installierte Komponenten
emr-5.31.1	5.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.31.0	5.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn



Amazon-EMR-Version sbezeichnung	Oozie-Version	Mit Oozie installierte Komponenten
emr-5.30.2	5.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.30.1	5.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Oozie-Version	Mit Oozie installierte Komponenten
emr-5.30.0	5.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.29.0	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Oozie-Version	Mit Oozie installierte Komponenten
emr-5.28.1	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.28.0	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Oozie-Version	Mit Oozie installierte Komponenten
emr-5.27.1	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.27.0	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Oozie-Version	Mit Oozie installierte Komponenten
emr-5.26.0	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.25.0	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Oozie-Version	Mit Oozie installierte Komponenten
emr-5.24.1	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.24.0	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Oozie-Version	Mit Oozie installierte Komponenten
emr-5.23.1	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.23.0	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Oozie-Version	Mit Oozie installierte Komponenten
emr-5.22.0	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.21.2	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn



Amazon-EMR-Version sbezeichnung	Oozie-Version	Mit Oozie installierte Komponenten
emr-5.21.1	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.21.0	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Oozie-Version	Mit Oozie installierte Komponenten
emr-5.20.1	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.20.0	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Oozie-Version	Mit Oozie installierte Komponenten
emr-5.19.1	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.19.0	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Oozie-Version	Mit Oozie installierte Komponenten
emr-5.18.1	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.18.0	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Oozie-Version	Mit Oozie installierte Komponenten
emr-5.17.2	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.17.1	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Oozie-Version	Mit Oozie installierte Komponenten
emr-5.17.0	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.16.1	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Oozie-Version	Mit Oozie installierte Komponenten
emr-5.16.0	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.15.1	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Oozie-Version	Mit Oozie installierte Komponenten
emr-5.15.0	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.14.2	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn



Amazon-EMR-Version sbezeichnung	Oozie-Version	Mit Oozie installierte Komponenten
emr-5.14.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.14.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Oozie-Version	Mit Oozie installierte Komponenten
emr-5.13.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.13.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Oozie-Version	Mit Oozie installierte Komponenten
emr-5.12.3	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.12.2	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Oozie-Version	Mit Oozie installierte Komponenten
emr-5.12.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.12.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Oozie-Version	Mit Oozie installierte Komponenten
emr-5.11.4	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.11.3	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Oozie-Version	Mit Oozie installierte Komponenten
emr-5.11.2	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.11.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Oozie-Version	Mit Oozie installierte Komponenten
emr-5.11.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.10.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Oozie-Version	Mit Oozie installierte Komponenten
emr-5.10.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.9.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn



Amazon-EMR-Version sbezeichnung	Oozie-Version	Mit Oozie installierte Komponenten
emr-5.9.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.8.3	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Oozie-Version	Mit Oozie installierte Komponenten
emr-5.8.2	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.8.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Oozie-Version	Mit Oozie installierte Komponenten
emr-5.8.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.7.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Oozie-Version	Mit Oozie installierte Komponenten
emr-5.7.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.6.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Oozie-Version	Mit Oozie installierte Komponenten
emr-5.6.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.5.4	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Oozie-Version	Mit Oozie installierte Komponenten
emr-5.5.3	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.5.2	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Oozie-Version	Mit Oozie installierte Komponenten
emr-5.5.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.5.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Oozie-Version	Mit Oozie installierte Komponenten
emr-5.4.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.4.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn



Amazon-EMR-Version sbezeichnung	Oozie-Version	Mit Oozie installierte Komponenten
emr-5.3.2	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.3.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Oozie-Version	Mit Oozie installierte Komponenten
emr-5.3.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.2.3	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Oozie-Version	Mit Oozie installierte Komponenten
emr-5.2.2	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.2.1	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Oozie-Version	Mit Oozie installierte Komponenten
emr-5.2.0	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.1.1	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Oozie-Version	Mit Oozie installierte Komponenten
emr-5.1.0	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.0.3	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Oozie-Version	Mit Oozie installierte Komponenten
emr-5.0.2	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.0.1	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon-EMR-Version sbezeichnung	Oozie-Version	Mit Oozie installierte Komponenten
emr-5.0.0	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

## Versionshinweise zu Oozie nach Version

### Themen

- [Amazon EMR 7.1.0 — Änderungen an Oozie](#)

### Amazon EMR 7.1.0 — Änderungen an Oozie

Typ	Beschreibung
Upgrade	Führt ein Upgrade von Netty auf 4.1.100.Final durch.
Upgrade	Führt ein Upgrade von Jetty auf 9.4.53.v20231009 durch.

# Apache Phoenix

Apache Phoenix wird für OLTP und betriebliche Analysen verwendet und ermöglicht Ihnen, SQL-Standardabfragen und JDBC-APIs für einen Apache HBase-Sicherheitsspeicher zu verwenden. Weitere Informationen finden Sie unter [Phoenix in 15 minutes or less](#). Phoenix ist in Amazon EMR Version 4.7.0 und höher enthalten.

Wenn Sie von einer früheren Version von Amazon EMR auf Amazon EMR Version 5.4.0 oder höher upgraden und die sekundäre Indizierung verwenden, aktualisieren Sie lokale Indizes wie in der [Apache Phoenix-Dokumentation](#) beschrieben. Amazon EMR entfernt die erforderlichen Konfigurationen aus der `hbase-site`-Klassifizierung, die Indizes müssen jedoch neu gefüllt werden. Online- und Offline-Upgrades von Indizes werden unterstützt. Online-Upgrades sind die Standardeinstellung. Dies bedeutet, dass Indizes neu gefüllt werden, während die Initialisierung von Phoenix-Clients Version 4.8.0 oder höher ausgeführt wird. Um Offline-Upgrades festzulegen, geben Sie für die `phoenix.client.localIndexUpgrade`-Konfiguration in der `phoenix-site`-Klassifizierung "false" und dann SSH für den Master-Knoten an, um `psql [zookeeper] -1` auszuführen.

Die folgende Tabelle listet die Version von Phoenix auf, die in der neuesten Version der Amazon-EMR-7.x-Serie enthalten ist, zusammen mit den Komponenten, die Amazon EMR mit Phoenix installiert.

Informationen zur Version der Komponenten, die in dieser Version mit Phoenix installiert wurden, finden Sie unter [Komponentenversionen von Version 7.1.0](#).

Phoenix-Versionsinformationen für emr-7.1.0

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-7.1.0	Phoenix 5.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager



Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
		nager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Die folgende Tabelle listet die Version von Phoenix auf, die in der neuesten Version der Amazon-EMR-6.x-Serie enthalten ist, zusammen mit den Komponenten, die Amazon EMR mit Phoenix installiert.

Die Version der Komponenten, die mit Phoenix in dieser Version installiert wurden, finden Sie unter [Komponentenversionen der Version 6.15.0](#).

#### Phoenix-Versionsinformationen für emr-6.15.0

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-6.15.0	Phoenix 5.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-operator-too

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
		ls, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Die folgende Tabelle listet die Version von Phoenix auf, die in der neuesten Version der Amazon-EMR-5.x-Serie enthalten ist, zusammen mit den Komponenten, die Amazon EMR mit Phoenix installiert.

Informationen zur Version der Komponenten, die in dieser Version mit Phoenix installiert wurden, finden Sie unter [Komponentenversionen von Version 5.36.2](#).

#### Phoenix-Versionsinformationen für emr-5.36.2

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-5.36.2	Phoenix 4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

## Themen

- [Einen Cluster mit Phoenix erstellen](#)
- [Phoenix-Clients](#)
- [Versionsverlauf von Phoenix](#)

## Einen Cluster mit Phoenix erstellen

Sie installieren Phoenix, indem Sie die Anwendung beim Erstellen eines Clusters in der Konsole oder über die AWS CLI auswählen. Die folgenden Verfahren und Beispiele zeigen, wie Sie einen Cluster mit Phoenix und HBase erstellen. Weitere Informationen zum Erstellen von Clustern mit der Konsole, einschließlich Erweiterte Optionen, finden Sie unter [Cluster planen und konfigurieren](#) im Verwaltungshandbuch für Amazon EMR.

So starten Sie einen Cluster mit installierter Phoenix-Anwendung mithilfe von Quick-Optionen zum Erstellen eines Clusters in der Konsole

1. Öffnen Sie die Amazon EMR-Konsole unter <https://console.aws.amazon.com/emr>.
2. Wählen Sie Create cluster (Cluster erstellen) aus, um Quick Create (Schnellerstellung) zu verwenden.
3. Wählen Sie bei Software Konfiguration die neueste Version für Ihre Anwendung aus. Phoenix wird nur als Option angezeigt, wenn Amazon Release Version emr-4.7.0 oder höher ausgewählt ist.
4. *Wählen Sie für **Anwendungen** die zweite Option, **HBase: HBase ver with Ganglia ver, Hadoop ver, Hive ver, Hue ver, Phoenix ver und ver. ZooKeeper***
5. Wählen Sie nach Bedarf weitere Optionen und anschließend Create cluster (Cluster erstellen) aus.

### Note

Linux-Zeilenfortsetzungszeichen (\) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (^).

Das folgende Beispiel startet einen Cluster mit installierter Phoenix-Anwendung mithilfe von Standard-Konfigurationseinstellungen.

Um einen Cluster mit Phoenix und HBase zu starten, verwenden Sie AWS CLI

- Erstellen Sie den Cluster mit dem folgenden Befehl:

```
aws emr create-cluster --name "Cluster with Phoenix" --release-label emr-7.1.0 \  
--applications Name=Phoenix Name=HBase --ec2-attributes KeyName=myKey \  
--instance-type m5.xlarge --instance-count 3 --use-default-roles
```

## Anpassen von Phoenix-Konfigurationen

Beim Erstellen eines Clusters konfigurieren Sie Phoenix, indem Sie Werte in `hbase-site.xml` mithilfe der `hbase-site`-Konfigurationsklassifizierung festlegen.

Weitere Informationen finden Sie unter [Konfiguration und Optimierung](#) in der Phoenix-Dokumentation.

Das folgende Beispiel zeigt, wie Sie mit einer in Amazon S3 gespeicherten JSON-Datei den Wert `false` für die Eigenschaft `phoenix.schema.dropMetaData` angeben. Für eine einzelne Klassifizierung können mehrere Eigenschaften angegeben werden. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#). Der Befehl `create-cluster` referenziert die JSON-Datei dann als Parameter `--configurations`.

Der Inhalt der JSON-Datei, die unter `/mybucket/myfolder/myconfig.json` gespeichert ist, lautet wie folgt.

```
[  
  {  
    "Classification": "hbase-site",  
    "Properties": {  
      "phoenix.schema.dropMetaData": "false"  
    }  
  }  
]
```

Der Befehl `create cluster`, der auf die JSON-Datei verweist, wird im folgenden Beispiel dargestellt.

```
aws emr create-cluster --release-label emr-7.1.0 --applications Name=Phoenix \  
--configurations s3://mybucket/myfolder/myconfig.json
```

```
Name=HBase --instance-type m5.xlarge --instance-count 2 \  
--configurations https://s3.amazonaws.com/mybucket/myfolder/myconfig.json
```

### Note

Rekonfigurationsanforderungen für Phoenix-Konfigurationsklassifizierungen werden nur in Amazon EMR ab Version 5.23.0, nicht jedoch in den Amazon-EMR-Versionen 5.21.0 und 5.22.0 unterstützt. Weitere Informationen finden Sie unter [Angeben einer Konfiguration für eine Instance-Gruppe in einem aktiven Cluster](#).

## Phoenix-Clients

Sie stellen eine Verbindung mit Phoenix entweder mithilfe eines JDBC-Clients mit vollständigen Abhängigkeiten oder mithilfe des "Thin Clients" her, der den Phoenix-Abfrageserver nutzt und nur auf einem Master-Knoten eines Clusters ausgeführt werden kann (z. B. durch die Verwendung eines SQL-Clients, eines Schritts, einer Befehlszeile, der SSH-Portweiterleitung usw.). Bei der Verwendung des "Fat" JDBC-Clients muss dennoch Zugriff auf alle Knoten des Clusters bestehen, da eine direkte Verbindung mit HBase-Services hergestellt wird. Der "Thin" Phoenix-Client muss nur auf den Phoenix-Abfrageserver über einen Standard-Port 8765 zugreifen. Es gibt mehrere [Skripts](#) in Phoenix, die diese Clients verwenden.

Einen Amazon-EMR-Schritt zum Abfragen mit Phoenix verwenden

Das folgende Verfahren stellt einen Snapshot von HBase wieder her und verwendet diese Daten zum Ausführen einer Phoenix-Abfrage. Sie können dieses Beispiel erweitern oder ein neues Skript erstellen, das die Phoenix-Clients an Ihre Anforderungen anpasst.

1. Erstellen Sie mit dem folgenden Befehl einen Cluster, auf dem Phoenix installiert ist:

```
aws emr create-cluster --name "Cluster with Phoenix" --log-uri s3://myBucket/  
myLogFolder --release-label emr-7.1.0 \  
--applications Name=Phoenix Name=HBase --ec2-attributes KeyName=myKey \  
--instance-type m5.xlarge --instance-count 3 --use-default-roles
```

2. Erstellen Sie die folgenden Dateien und laden Sie sie in Amazon S3: hoch:

copySnapshot.sh

```
sudo su hbase -s /bin/sh -c 'hbase snapshot export \
-D hbase.rootdir=s3://us-east-1.elasticmapreduce.samples/hbase-demo-customer-data/
snapshot/ \
-snapshot customer_snapshot1 \
-copy-to hdfs://masterDNSName:8020/user/hbase \
-mappers 2 -chuser hbase -chmod 700'
```

## runQuery.sh

```
aws s3 cp s3://myBucket/phoenixQuery.sql /home/hadoop/
/usr/lib/phoenix/bin/sqlline-thin.py http://localhost:8765 /home/hadoop/
phoenixQuery.sql
```

## phoenixQuery.sql

### Note

Sie müssen COLUMN\_ENCODED\_BYTES=0 in das folgende Beispiel nur einbeziehen, wenn Sie Amazon-EMR-Versionen 5.26.0 und höher verwenden.

```
CREATE VIEW "customer" (
pk VARCHAR PRIMARY KEY,
"address"."state" VARCHAR,
"address"."street" VARCHAR,
"address"."city" VARCHAR,
"address"."zip" VARCHAR,
"cc"."number" VARCHAR,
"cc"."expire" VARCHAR,
"cc"."type" VARCHAR,
"contact"."phone" VARCHAR)
COLUMN_ENCODED_BYTES=0;

CREATE INDEX my_index ON "customer" ("customer"."state") INCLUDE("PK",
"customer"."city", "customer"."expire", "customer"."type");

SELECT "customer"."type" AS credit_card_type, count(*) AS num_customers FROM
"customer" WHERE "customer"."state" = 'CA' GROUP BY "customer"."type";
```

Verwenden Sie den AWS CLI , um die Dateien an den S3-Bucket zu senden:

```
aws s3 cp copySnapshot.sh s3://myBucket/
aws s3 cp runQuery.sh s3://myBucket/
aws s3 cp phoenixQuery.sql s3://myBucket/
```

- Erstellen Sie eine Tabelle mithilfe des folgenden, an den Cluster gesendeten Schritts, den Sie in Schritt 1 erstellt haben:

createTable.json

```
[
  {
    "Name": "Create HBase Table",
    "Args": ["bash", "-c", "echo '$'create \"customer\", \"address\", \"cc\", \"contact\" | hbase shell"],
    "Jar": "command-runner.jar",
    "ActionOnFailure": "CONTINUE",
    "Type": "CUSTOM_JAR"
  }
]
```

```
aws emr add-steps --cluster-id j-2AXXXXXXGAPLF \
--steps file:///./createTable.json
```

- Verwenden Sie `script-runner.jar`, um das zuvor in Ihren S3-Bucket hochgeladene `copySnapshot.sh`-Skript auszuführen:

```
aws emr add-steps --cluster-id j-2AXXXXXXGAPLF \
--steps Type=CUSTOM_JAR,Name="HBase Copy Snapshot",ActionOnFailure=CONTINUE,\
Jar=s3://region.elasticmapreduce/libs/script-runner/script-
runner.jar,Args=["s3://myBucket/copySnapshot.sh"]
```

Dadurch wird ein MapReduce Job ausgeführt, um Ihre Snapshot-Daten auf das Cluster-HDFS zu kopieren.

- Stellen Sie den Snapshot, den Sie in den Cluster kopiert haben, mit dem folgenden Schritt wieder her:

restoreSnapshot.json

```
[
  {
    "Name": "restore",
    "Args": ["bash", "-c", "echo '$disable \"customer\"; restore_snapshot \"customer_snapshot1\"; enable \"customer\"' | hbase shell"],
    "Jar": "command-runner.jar",
    "ActionOnFailure": "CONTINUE",
    "Type": "CUSTOM_JAR"
  }
]
```

```
aws emr add-steps --cluster-id j-2AXXXXXXGAPLF \
--steps file:///./restoreSnapshot.json
```

6. Verwenden Sie `script-runner.jar`, um das zuvor in Ihren S3-Bucket hochgeladene `runQuery.sh`-Skript auszuführen:

```
aws emr add-steps --cluster-id j-2AXXXXXXGAPLF \
--steps Type=CUSTOM_JAR,Name="Phoenix Run Query",ActionOnFailure=CONTINUE,\
Jar=s3://region.elasticmapreduce/libs/script-runner/script-
runner.jar,Args=["s3://myBucket/runQuery.sh"]
```

Die Abfrage wird ausgeführt und gibt das Ergebnis an `stdout` des Schritts zurück. Es kann einige Minuten dauern, bis dieser Schritt abgeschlossen ist.

7. Prüfen Sie die Ergebnisse von `stdout` des Schritts an der Protokoll-URL, mit der Sie den Cluster in Schritt 1 erstellt haben. Die Ergebnisse sollten wie folgt aussehen:

```
+-----+-----+
|          CREDIT_CARD_TYPE          |          NUM_CUSTOMERS          |
+-----+-----+
| american_express                    | 5728                            |
| dankort                             | 5782                            |
| diners_club                         | 5795                            |
| discover                            | 5715                            |
| forbrugsforeningen                 | 5691                            |
| jcb                                 | 5762                            |
| laser                               | 5769                            |
| maestro                             | 5816                            |
| mastercard                         | 5697                            |
| solo                                | 5586                            |
+-----+-----+
```



switch	5781	
visa	5659	
+-----+	+-----+	+-----+

## Versionsverlauf von Phoenix

In der folgenden Tabelle sind die Version von Phoenix aufgeführt, die in jeder Release-Version von Amazon EMR enthalten ist, zusammen mit den Komponenten, die mit der Anwendung installiert wurden. Informationen zu den Komponentenversionen in den einzelnen Versionen finden Sie im Abschnitt Komponentenversion für Ihre Version in [Amazon-EMR-7.x-Versionen](#), [Amazon-EMR-6.x-Versionen](#) oder [Amazon-EMR-5.x-Versionen](#).

### Versionsinformationen für Phoenix

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-5.36.2	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-7.1.0	5.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
		library, hadoop-hdfs-nameno de, hadoop-https-server, hadoop-kms-server, hadoop- mapred, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hbase-hma ster, hbase-client, hbase-reg ion-server, hbase-operator-too ls, phoenix-library, phoenix- connectors, phoenix-query- server, zookeeper-client, zookeeper-server
emr-7.0.0	5.1.3	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop- hdfs-datanode, hadoop-hdfs- library, hadoop-hdfs-nameno de, hadoop-https-server, hadoop-kms-server, hadoop- mapred, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hbase-hma ster, hbase-client, hbase-reg ion-server, hbase-operator-too ls, phoenix-library, phoenix- connectors, phoenix-query- server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-6.15.0	5.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-6.14.0	5.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-6.13.0	5.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-6.12.0	5.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-6.11.1	5.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-6.11.0	5.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server



Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-6.10.1	5.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-6.10.0	5.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-6.9.1	5.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-6.9.0	5.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-6.8.1	5.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-6.8.0	5.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-6.7.0	5.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-5.36.1	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.36.0	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server



Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-6.6.0	5.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-5.35.0	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-6.5.0	5.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-6.4.0	5.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-6.3.1	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-6.3.0	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-6.2.1	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-6.2.0	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-6.1.1	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-6.1.0	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-6.0.1	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-6.0.0	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.34.0	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-5.33.1	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.33.0	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server



Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-5.32.1	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.32.0	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-5.31.1	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.31.0	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-5.30.2	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.30.1	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-5.30.0	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.29.0	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-5.28.1	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.28.0	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-5.27.1	4.14.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.27.0	4.14.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-5.26.0	4.14.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.25.0	4.14.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-5.24.1	4.14.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.24.0	4.14.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server



Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-5.23.1	4.14.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.23.0	4.14.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-5.22.0	4.14.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.21.2	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-5.21.1	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.21.0	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-5.20.1	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.20.0	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-5.19.1	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.19.0	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-5.18.1	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.18.0	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-5.17.2	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.17.1	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-5.17.0	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.16.1	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server



Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-5.16.0	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.15.1	4.13.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-5.15.0	4.13.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.14.2	4.13.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-5.14.1	4.13.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.14.0	4.13.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-5.13.1	4.13.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.13.0	4.13.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-5.12.3	4.13.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.12.2	4.13.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-5.12.1	4.13.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.12.0	4.13.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-5.11.4	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.11.3	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-5.11.2	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.11.1	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server



Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-5.11.0	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.10.1	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-5.10.0	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.9.1	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-5.9.0	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.8.3	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-5.8.2	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.8.1	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-5.8.0	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.7.1	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-5.7.0	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.6.1	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-5.6.0	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.5.4	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-5.5.3	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpserver, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.5.2	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpserver, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server



Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-5.5.1	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpserver, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.5.0	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpserver, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-5.4.1	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.4.0	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-5.3.2	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpserver, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.3.1	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpserver, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-5.3.0	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.2.3	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-5.2.2	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpserver, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.2.1	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpserver, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-5.2.0	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpserver, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.1.1	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpserver, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-5.1.0	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpserver, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.0.3	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpserver, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-5.0.2	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpserver, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.0.1	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpserver, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server



Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-5.0.0	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpserver, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-4.9.6	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpserver, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-4.9.5	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpserver, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-4.9.4	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpserver, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-4.9.3	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpserver, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-4.9.2	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpserver, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-4.9.1	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpserver, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-4.8.5	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpserver, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-4.8.4	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpserver, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-4.8.3	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpserver, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-4.8.2	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpserver, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-4.8.1	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpserver, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-4.8.0	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpserver, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-4.7.4	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpserver, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-4.7.3	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpserver, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-4.7.2	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpserver, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server



Amazon-EMR-Version sbezeichnung	Phoenix-Version	Mit Phoenix installierte Komponenten
emr-4.7.1	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpserver, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-4.7.0	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpserver, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

# Apache Pig

Apache Pig ist eine Open-Source-Apache-Bibliothek, die auf Hadoop basiert und eine Skriptsprache bereitstellt, die Sie verwenden können, um große Datasets zu transformieren, ohne dass Sie komplexen Code in einer Computersprache einer niedrigeren Ebene schreiben müssen, wie etwa Java. Die Bibliothek verwendet SQL-ähnliche Befehle, die in einer Sprache namens Pig Latin geschrieben sind, und konvertiert diese Befehle in Tez-Jobs, die auf gerichteten azyklischen Graphen (DAGs) oder Programmen basieren. MapReduce Pig funktioniert mit strukturierten und unstrukturierten Daten in verschiedenen Formaten. Weitere Informationen zu Pig finden Sie unter <http://pig.apache.org/>.

Sie können Pig-Befehle interaktiv oder im Batch-Modus ausführen. Um Pig interaktiv zu nutzen, erstellen Sie eine SSH-Verbindung mit dem Master-Knoten und senden Befehle per Grunt-Shell. Um Pig im Batch-Modus zu nutzen, laden Sie Ihre Pig-Skripts in Amazon S3 hoch und senden diese dann als Cluster-Schritte. Weitere Informationen zum Übermitteln von Aufträgen an einen Cluster finden Sie unter [Aufträge an einen Cluster übermitteln](#) im Verwaltungshandbuch für Amazon EMR.

Wenn Sie Pig verwenden, um Ausgaben in eine HCatalog-Tabelle in Amazon S3 zu schreiben, deaktivieren Sie Amazon EMR Direct Write, indem Sie die Eigenschaften `mapred.output.direct.NativeS3FileSystem` und `mapred.output.direct.EmrFileSystem` auf `false` setzen. Weitere Informationen finden Sie unter [Verwenden von HCatalog](#). Innerhalb eines Pig-Skripts können Sie die Befehle `SET mapred.output.direct.NativeS3FileSystem false` und `SET mapred.output.direct.EmrFileSystem false` verwenden.

Die folgende Tabelle listet die Version von Pig auf, die in der neuesten Version der Amazon-EMR-7.x-Serie enthalten ist, zusammen mit den Komponenten, die Amazon EMR mit Pig installiert.

[Die Version der Komponenten, die in dieser Version zusammen mit Pig installiert wurden, finden Sie unter Komponentenversionen von Version 7.1.0.](#)

Pig-Versionsinformationen für emr-7.1.0

Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponenten
emr-7.1.0	Pig 0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-

Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponenten
		cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn, tez-on-worker

Die folgende Tabelle listet die Version von Pig auf, die in der neuesten Version der Amazon-EMR-6.x-Serie enthalten ist, zusammen mit den Komponenten, die Amazon EMR mit Pig installiert.

Die Version der Komponenten, die mit Pig in dieser Version installiert wurden, finden Sie unter [Komponentenversionen der Version 6.15.0](#).

#### Pig-Versionsinformationen für emr-6.15.0

Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponenten
emr-6.15.0	Pig 0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn, tez-on-worker

Die folgende Tabelle listet die Version von Pig auf, die in der neuesten Version der Amazon-EMR-5.x-Serie enthalten ist, zusammen mit den Komponenten, die Amazon EMR mit Pig installiert.

Informationen zur Version der Komponenten, die in dieser Version mit Pig installiert wurden, finden Sie unter [Komponentenversionen von Version 5.36.2](#).

Pig-Versionsinformationen für emr-5.36.2

Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponenten
emr-5.36.2	Pig 0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Themen

- [Übermitteln von Pig-Aufträgen](#)
- [Benutzerdefinierte Funktionen in Pig aufrufen](#)
- [Pig-Versionsverlauf](#)

## Übermitteln von Pig-Aufträgen

Dieser Abschnitt zeigt das Senden von Pig-Aufträgen an einen Amazon-EMR-Cluster. Das folgende Beispiel generiert einen Bericht mit der Gesamtzahl der übertragenen Bytes, eine Liste der Top 50 IP-Adressen, eine Liste der externen Top 50-Referrer und die 50 beliebtesten Suchbegriffe bei Bing und Google. Das Pig-Skript befindet sich im Amazon-S3-Bucket `s3://elasticmapreduce/samples/pig-apache/do-reports2.pig`. Die Eingabedaten befinden sich im Amazon-S3-Bucket `s3://`

elasticmapreduce/samples/pig-apache/input. Die Ausgabe wird in einem Amazon-S3-Bucket gespeichert.

## Pig-Aufträge über die Amazon-EMR-Konsole senden

In diesem Beispiel wird beschrieben, wie Sie mithilfe der Amazon-EMR-Konsole einen Pig-Auftrag an einen Cluster senden.

So senden Sie einen Pig-Auftrag

1. Öffnen Sie die Amazon EMR-Konsole unter <https://console.aws.amazon.com/emr>.
2. Wählen Sie Cluster erstellen, um einen neuen Cluster mit Pig zu erstellen. Anweisungen zum Erstellen eines Clusters finden Sie unter [Planen und Konfigurieren eines Amazon-EMR-Clusters](#).
3. Öffnen Sie ein Terminal und stellen Sie eine SSH-Verbindung zum Hauptknoten Ihres Clusters her, indem Sie die unter [Verbindung zum Hauptknoten mithilfe von SSH](#) hergestellten beschriebenen Schritte ausführen. Sobald Sie das getan haben, führen Sie die folgenden Schritte aus.

```
sudo mkdir -p /home/hadoop/lib/pig/  
sudo aws s3 cp s3://elasticmapreduce/libs/pig/0.3/piggybank-0.3-amzn.jar /home/  
hadoop/lib/pig/piggybank.jar
```

4. Klicken Sie in der Konsole die Seite Cluster-Liste aus und klicken Sie den Namen des von Ihnen erstellten Clusters.
5. Scrollen Sie zum Abschnitt Steps (Schritte) und erweitern Sie ihn. Wählen Sie anschließend Add step (Schritt hinzufügen) aus.
6. Gehen Sie im Dialogfeld Schritt hinzufügen folgendermaßen vor:
  - Wählen Sie unter Step type (Schritttyp) die Option Pig program (Pig-Programm) aus.
  - Übernehmen Sie unter Name den Standardnamen („Pig program“) oder geben Sie einen neuen Namen ein.
  - Geben Sie unter Script S3 location (S3-Skriptspeicherort) den Speicherort Ihres Pig-Skripts ein. Zum Beispiel: **s3://elasticmapreduce/samples/pig-apache/reports2.pig**.
  - Unter Input S3 location (S3-Eingabespeicherort) geben Sie den Speicherort der Eingabedaten ein. Zum Beispiel: **s3://elasticmapreduce/samples/pig-apache/input**.

- Unter S3-Ausgabespeicherort geben Sie den Namen Ihres Amazon-S3-Ausgabe-Buckets ein oder navigieren zu dem entsprechenden S3-Bucket.
  - Lassen Sie unter Arguments (Parameter) das Feld leer.
  - Übernehmen Sie unter Action on failure (Aktion bei Fehler) die Standardeinstellung Continue (Fortfahren).
7. Wählen Sie Hinzufügen aus. Der Schritt wird in der Konsole mit dem Status "Pending" angezeigt.
  8. Bei Ausführung des Schritts ändert sich dessen Status von Pending zu Running und Completed. Wählen Sie über der Spalte Actions (Aktionen) das Symbol Refresh (Ansicht aktualisieren) aus, um den Status zu aktualisieren. Wenn Ihr Schritt abgeschlossen ist, überprüfen Sie Ihren Amazon-S3-Bucket, um sicherzustellen, dass die Ausgabedateien Ihres Pig-Schritts vorhanden sind.

## Reichen Sie Arbeiten von Pig ein, indem Sie AWS CLI

Um einen Pig-Schritt einzureichen, verwenden Sie den AWS CLI

Wenn Sie einen Cluster mit dem starten AWS CLI, verwenden Sie den `--applications` Parameter, um Pig zu installieren. Verwenden Sie den Parameter `--steps`, um einen Pig-Schritt zu übermitteln.

1. Um einen Cluster mit installiertem Pig zu starten, geben Sie den folgenden Befehl ein und ersetzen Sie *myKey* und *DOC-EXAMPLE-BUCKET/* durch den Namen Ihres EC2-Schlüsselpaars und Ihres Amazon-S3-Buckets.

```
aws emr create-cluster \  
--name "Test cluster" \  
--log-uri s3://DOC-EXAMPLE-BUCKET/ \  
--release-label emr-5.36.2 \  
--applications Name=Pig \  
--use-default-roles \  
--ec2-attributes KeyName=myKey \  
--instance-type m5.xlarge \  
--instance-count 3
```

**Note**

Linux-Zeilenfortsetzungszeichen (`\`) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (`^`).

Wenn Sie die Instance-Anzahl ohne den Parameter `--instance-groups` angeben, wird ein einzelner Master-Knoten gestartet. Die verbleibenden Instances werden als Core-Knoten aufgerufen. Alle Knoten verwenden den im Befehl angegebenen Instance-Typ.

**Note**

Wenn Sie zuvor nicht die standardmäßige EMR-Servicerolle und das EC2-Instance-Profil erstellt haben, geben Sie `aws emr create-default-roles` ein, um sie zu erstellen, bevor Sie den Unterbefehl `create-cluster` eingeben.

- Um einen Pig-Schritt einzureichen, geben Sie den folgenden Befehl ein *myClusterId* und ersetzen Sie *DOC-EXAMPLE-BUCKET* durch Ihre Cluster-ID und den Namen Ihres Amazon S3 S3-Buckets.

```
aws emr add-steps \  
--cluster-id myClusterId \  
--steps Type=PIG,Name="Pig Program",ActionOnFailure=CONTINUE,Args=[-f,s3://  
elasticmapreduce/samples/pig-apache/do-reports2.pig,-p,INPUT=s3://elasticmapreduce/  
samples/pig-apache/input,-p,OUTPUT=s3://DOC-EXAMPLE-BUCKET/pig-apache/output]
```

Dieser Befehl gibt eine Schritt-ID zurück, anhand derer Sie die Ausführung Ihres State-Schritts überprüfen können.

- Fragen Sie den Status Ihres Schritts mit dem `describe-step`-Befehl ab.

```
aws emr describe-step --cluster-id myClusterId --step-id s-1XXXXXXXXXXA
```

Der State-Wert des Schritts ändert sich mit der Ausführung des Schritts von PENDING zu RUNNING zu COMPLETED. Wenn Ihr Schritt abgeschlossen ist, überprüfen Sie Ihren Amazon-S3-Bucket, um sicherzustellen, dass die Ausgabedateien Ihres Pig-Schritts vorhanden sind.

Weitere Informationen zur Verwendung von Amazon EMR-Befehlen in der AWS CLI finden Sie in der [AWS CLI Befehlsreferenz](#).

## Benutzerdefinierte Funktionen in Pig aufrufen

Pig bietet die Möglichkeit zum Aufrufen von benutzerdefinierten Funktionen (User Defined Functions, UDFs) in Pig-Skripts. Sie können so eine benutzerdefinierte Verarbeitung in Ihren Pig-Skripts implementieren. Die derzeit unterstützten Sprachen sind Java, Python/Jython und JavaScript (obwohl die JavaScript Unterstützung noch experimentell ist.)

In den folgenden Abschnitten wird beschrieben, wie Sie Ihre Funktionen registrieren, um diese über die Pig-Shell oder in Pig-Skripts aufzurufen. Weitere Informationen zur Verwendung von UDFs mit Pig finden Sie in der [Pig-Dokumentation](#) für Ihre Pig-Version.

### JAR-Dateien in Pig aufrufen

Mit dem Befehl REGISTER können Sie in Ihren Pig-Skripts benutzerdefinierte JAR-Dateien nutzen. Die JAR-Datei ist im lokalen Dateisystem oder in einem Remote-Dateisystem z. B. Amazon S3 gespeichert. Wenn das Pig-Skript ausgeführt wird, lädt Amazon EMR die JAR-Datei automatisch auf den Hauptknoten herunter und diese dann in den verteilten Hadoop-Cache hoch. So wird die JAR-Datei bei Bedarf automatisch von allen Instances im Cluster verwendet.

So verwenden Sie JAR-Dateien mit Pig

1. Laden Sie Ihre benutzerdefinierte JAR-Datei in Amazon S3 hoch.
2. Mit dem Befehl REGISTER in Ihren Pig-Skripts geben Sie den Amazon-S3-Bucket mit der benutzerdefinierten JAR-Datei an.

```
REGISTER s3://mybucket/path/mycustomjar.jar;
```

### Aufrufen von Python/Jython-Skripts in Pig

Sie können Python-Skripts in Pig registrieren und diese Skripts dann über die Pig-Shell oder in einem Pig-Skript aufrufen. Dies geschieht, indem Sie den Speicherort des Skripts mit dem Schlüsselwort `register` angeben.

Da Pig in Java geschrieben ist, verwendet es die Jython-Skript-Engine zum Parsen des Python-Skripts. Weitere Informationen zu Jython finden Sie unter <http://www.jython.org/>.



## So rufen Sie ein Python/Jython-Skript in Pig auf

1. Schreiben Sie ein Python-Skript und laden Sie es an einem Speicherort in Amazon S3 hoch. Es sollte sich um einen Bucket handeln, der sich im Besitz des Kontos befindet, das den Pig-Cluster erstellt bzw. das über entsprechende Berechtigungen verfügt. Nur so kann das Konto auf den Cluster zugreifen. In diesem Beispiel wird das Skript auf `s3://mybucket/pig/python` hochgeladen.
2. Starten Sie einen Pig-Cluster. Wenn Sie über die Grunt-Shell auf Pig zugreifen, starten Sie einen interaktiven Cluster. Wenn Sie Pig-Befehle aus einem Skript heraus ausführen, starten Sie einen skriptgesteuerten Pig-Cluster. In diesem Beispiel wird ein interaktiver Cluster gestartet. Weitere Informationen zum Erstellen eines Pig-Clusters finden Sie unter [Übermitteln von Pig-Aufträgen](#).
3. Verwenden Sie bei einem interaktiven Cluster SSH für die Verbindung mit dem Master-Knoten und führen Sie die Grunt-Shell aus. Weitere Informationen finden Sie unter [SSH im Hauptknoten](#).
4. Führen Sie die Grunt-Shell für Pig aus, indem Sie `pig` in der Befehlszeile eingeben:

```
pig
```

5. Registrieren Sie mit dem Schlüsselwort `register` in der Grunt-Eingabeaufforderung die Jython-Bibliothek und Ihr Python-Skript in Pig (wie im folgenden Beispiel gezeigt). Geben Sie den Speicherort Ihres Skripts in Amazon S3 an:

```
grunt> register 'lib/jython.jar';  
grunt> register 's3://mybucket/pig/python/myscript.py' using jython as myfunctions;
```

6. Laden Sie die Eingabedaten. Beim folgenden Beispiel wird die Eingabe aus einem Amazon-S3-Speicherort geladen:

```
grunt> input = load 's3://mybucket/input/data.txt' using TextLoader as  
(line:chararray);
```

7. Indem Sie mit `myfunctions` auf sie verweisen, können Sie jetzt Funktionen in Ihrem Pig-Skript aufrufen:

```
grunt> output=foreach input generate myfunctions.myfunction($1);
```

## Pig-Versionsverlauf

In der folgenden Tabelle sind die Version von Pig aufgeführt, die in jeder Release-Version von Amazon EMR enthalten ist, zusammen mit den Komponenten, die mit der Anwendung installiert wurden. Informationen zu den Komponentenversionen in den einzelnen Versionen finden Sie im Abschnitt [Komponentenversion für Ihre Version in Amazon-EMR-7.x-Versionen](#), [Amazon-EMR-6.x-Versionen](#) oder [Amazon-EMR-5.x-Versionen](#).

Die Pig Versionsinformationen

Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponent en
emr-5.36.2	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-7.1.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn, tez-on-worker

Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponent en
emr-7.0.0	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn, tez-on-worker
emr-6.15.0	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn, tez-on-worker

Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponent en
emr-6.14.0	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn, tez-on-worker
emr-6.13.0	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn, tez-on-worker

Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponent en
emr-6.12.0	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn, tez-on-worker
emr-6.11.1	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn, tez-on-worker

Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponent en
emr-6.11.0	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn, tez-on-worker
emr-6.10.1	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn, tez-on-worker

Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponent en
emr-6.10.0	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn, tez-on-worker
emr-6.9.1	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn

Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponent en
emr-6.9.0	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn
emr-6.8.1	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn



Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponent en
emr-6.8.0	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn
emr-6.7.0	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn

Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponent en
emr-5.36.1	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn
emr-5.36.0	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn

Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponent en
emr-6.6.0	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn
emr-5.35.0	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn

Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponent en
emr-6.5.0	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn
emr-6.4.0	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn

Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponent en
emr-6.3.1	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn
emr-6.3.0	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn

Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponent en
emr-6.2.1	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn
emr-6.2.0	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn

Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponent en
emr-6.1.1	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn
emr-6.1.0	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn

Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponent en
emr-5.34.0	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn
emr-5.33.1	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn



Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponent en
emr-5.33.0	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn
emr-5.32.1	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn

Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponent en
emr-5.32.0	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn
emr-5.31.1	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn

Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponent en
emr-5.31.0	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn
emr-5.30.2	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn

Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponent en
emr-5.30.1	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn
emr-5.30.0	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn

Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponent en
emr-5.29.0	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn
emr-5.28.1	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn

Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponent en
emr-5.28.0	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn
emr-5.27.1	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn

Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponent en
emr-5.27.0	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn
emr-5.26.0	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn

Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponent en
emr-5.25.0	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn
emr-5.24.1	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn



Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponent en
emr-5.24.0	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn
emr-5.23.1	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn

Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponent en
emr-5.23.0	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn
emr-5.22.0	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn

Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponent en
emr-5.21.2	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn
emr-5.21.1	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn

Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponent en
emr-5.21.0	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn
emr-5.20.1	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn

Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponent en
emr-5.20.0	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn
emr-5.19.1	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn

Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponent en
emr-5.19.0	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn
emr-5.18.1	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn

Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponent en
emr-5.18.0	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn
emr-5.17.2	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn

Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponent en
emr-5.17.1	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn
emr-5.17.0	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn



Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponent en
emr-5.16.1	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn
emr-5.16.0	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn

Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponent en
emr-5.15.1	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn
emr-5.15.0	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn

Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponent en
emr-5.14.2	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn
emr-5.14.1	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn

Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponent en
emr-5.14.0	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn
emr-5.13.1	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn

Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponent en
emr-5.13.0	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn
emr-5.12.3	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn

Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponent en
emr-5.12.2	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn
emr-5.12.1	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn

Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponent en
emr-5.12.0	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn
emr-5.11.4	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn

Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponent en
emr-5.11.3	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn
emr-5.11.2	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn



Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponent en
emr-5.11.1	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn
emr-5.11.0	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn

Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponent en
emr-5.10.1	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn
emr-5.10.0	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn

Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponent en
emr-5.9.1	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn
emr-5.9.0	0.17.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn

Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponent en
emr-5.8.3	0.16.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn
emr-5.8.2	0,16.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn

Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponent en
emr-5.8.1	0,16.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn
emr-5.8.0	0,16.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn

Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponent en
emr-5.7.1	0,16.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn
emr-5.7.0	0,16.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn

Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponent en
emr-5.6.1	0,16.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn
emr-5.6.0	0,16.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn

Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponent en
emr-5.5.4	0,16.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn
emr-5.5.3	0,16.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn



Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponent en
emr-5.5.2	0,16.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn
emr-5.5.1	0,16.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn

Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponent en
emr-5.5.0	0,16.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn
emr-5.4.1	0,16.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn

Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponent en
emr-5.4.0	0,16.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn
emr-5.3.2	0,16.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn

Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponent en
emr-5.3.1	0,16.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn
emr-5.3.0	0,16.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn

Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponent en
emr-5.2.3	0,16.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn
emr-5.2.2	0,16.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn

Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponent en
emr-5.2.1	0,16.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn
emr-5.2.0	0,16.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn

Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponent en
emr-5.1.1	0,16.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn
emr-5.1.0	0,16.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn

Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponent en
emr-5.0.3	0,16.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn
emr-5.0.2	0,16.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn



Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponent en
emr-5.0.1	0,16.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn
emr-5.0.0	0,16.0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, pig-client, tez- on-yarn

Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponent en
emr-4.9.6	0,14,0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, pig-client
emr-4.9.5	0,14,0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, pig-client
emr-4.9.4	0,14,0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, pig-client

Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponent en
emr-4.9.3	0,14,0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, pig-client
emr-4.9.2	0,14,0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, pig-client
emr-4.9.1	0,14,0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, pig-client

Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponent en
emr-4.8.5	0,14,0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, pig-client
emr-4.8.4	0,14,0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, pig-client
emr-4.8.3	0,14,0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, pig-client

Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponent en
emr-4.8.2	0,14,0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, pig-client
emr-4.8.1	0,14,0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, pig-client
emr-4.8.0	0,14,0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, pig-client

Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponent en
emr-4.7.4	0,14,0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, pig-client
emr-4.7.3	0,14,0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, pig-client
emr-4.7.2	0,14,0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, pig-client

Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponent en
emr-4.7.1	0,14,0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, pig-client
emr-4.7.0	0,14,0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, pig-client
emr-4.6.1	0,14,0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-ht tpfs-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, pig-client

Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponent en
emr-4.6.0	0,14,0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-ht tpfs-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, pig-client
emr-4.5.0	0,14,0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-ht tpfs-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, pig-client
emr-4.4.0	0,14,0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-ht tpfs-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, pig-client



Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponent en
emr-4.3.0	0,14,0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-ht tpfs-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, pig-client
emr-4.2.0	0,14,0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-ht tpfs-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, pig-client
emr-4.1.0	0,14,0	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop-ma pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop-ht tpfs-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, pig-client

Amazon-EMR-Version sbezeichnung	Pig-Version	Mit Pig installierte Komponenten
emr-4.0.0	0,14,0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datano-de, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client

# Presto und Trino

## Note

PrestoSQL wurde im Dezember 2020 in Trino umbenannt. Amazon-EMR-Versionen 6.4.0 und höher verwenden den Namen Trino, während frühere Versionen den Namen PrestoSQL verwenden.

[Presto](#) ist eine schnelle SQL-Abfrage-Engine für interaktive Analyseabfragen für große Datensätze aus verschiedenen Quellen. Weitere Informationen finden Sie auf der [Presto-Website](#). Presto ist in Amazon-EMR-Versionen 5.0.0 und höher enthalten. Frühere Versionen enthalten Presto als Sandbox-Anwendung. Weitere Informationen finden Sie unter [Amazon-EMR-4.x-Versionen](#). Die Amazon-EMR-Versionen 6.1.0 und höher unterstützen zusätzlich zu Presto auch [Trino](#) (PrestoSQL). Weitere Informationen finden Sie unter [Installation von PrestoDB und Trino](#).

Die folgende Tabelle listet die Version von Presto auf, die in der neuesten Version der Amazon-EMR-7.x-Serie enthalten ist, zusammen mit den Komponenten, die Amazon EMR mit Presto installiert.

Informationen zur Version der Komponenten, die in dieser Version mit Presto installiert wurden, finden Sie unter [Komponentenversionen von Version 7.1.0](#).

## Presto-Versionsinformationen für emr-7.1.0

Amazon-EMR-Version sbezeichnung	Presto-Version	Mit Presto installierte Komponenten
emr-7.1.0	Presto 0.284	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-

Amazon-EMR-Version sbezeichnung	Presto-Version	Mit Presto installierte Komponenten
		server, presto-coordinator, presto-worker

Die folgende Tabelle listet die Version von Presto auf, die in der neuesten Version der Amazon-EMR-6.x-Serie enthalten ist, zusammen mit den Komponenten, die Amazon EMR mit Presto installiert.

Die Version der Komponenten, die mit Presto in dieser Version installiert wurden, finden Sie unter [Komponentenversionen der Version 6.15.0](#).

Presto-Versionsinformationen für emr-6.15.0

Amazon-EMR-Version sbezeichnung	Presto-Version	Mit Presto installierte Komponenten
emr-6.15.0	Presto 0.283	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Die folgende Tabelle listet die Version von Presto auf, die in der neuesten Version der Amazon-EMR-5.x-Serie enthalten ist, zusammen mit den Komponenten, die Amazon EMR mit Presto installiert.

[Informationen zur Version der Komponenten, die in dieser Version mit Presto installiert wurden, finden Sie unter Komponentenversionen von Version 5.36.2.](#)

## Presto-Versionsinformationen für emr-5.36.2

Amazon-EMR-Version sbezeichnung	Presto-Version	Mit Presto installierte Komponenten
emr-5.36.2	Presto 0.267	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Die folgende Tabelle listet die Version von Trino (PrestoSQL) auf, die in der neuesten Version der Amazon-EMR-6.x-Serie enthalten ist, zusammen mit den Komponenten, die Amazon EMR mit Trino (PrestoSQL) installiert.

Die Version der Komponenten, die mit Trino (PrestoSQL) in dieser Version installiert wurden, finden Sie unter [Komponentenversionen der Version 6.15.0](#).

## Trino (PrestoSQL)-Versionsinformationen für emr-6.15.0

Amazon-EMR-Version sbezeichnung	Trino (PrestoSQL)-Version	Mit Trino (PrestoSQL) installierte Komponenten
emr-6.15.0	Trino (PrestoSQL) 426	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client,

Amazon-EMR-Version sbezeichnung	Trino (PrestoSQL)-Version	Mit Trino (PrestoSQL) installierte Komponenten
		hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker

## Themen

- [Presto mit dem AWS Glue-Datenkatalog verwenden](#)
- [S3-Select-Pushdown mit Presto zur Verbesserung der Leistung verwenden](#)
- [Hinzufügen von Datenbank-Konnektoren](#)
- [Verwenden von SSL/TLS und Konfigurieren von LDAPS mit Presto in Amazon EMR](#)
- [Den strikten Presto-Modus aktivieren](#)
- [Umgang mit dem Verlust von Spot Instances in Presto](#)
- [Fehlertolerante Ausführung in Trino](#)
- [Auto Scaling von Presto mit ordnungsgemäßer Stilllegung verwenden](#)
- [Überlegungen zu Presto in Amazon EMR](#)
- [Presto-Versionsverlauf](#)

## Presto mit dem AWS Glue-Datenkatalog verwenden

Mit der Amazon EMR-Version 5.10.0 und höher können Sie den AWS Glue-Datenkatalog als Standard-Hive-Metastore für Presto angeben. Wir empfehlen diese Konfiguration, wenn Sie einen persistenten Metastore oder einen gemeinsamen Metastore für verschiedene Cluster, Services, Anwendungen oder AWS-Konten benötigen.

AWS Glue ist ein vollständig verwalteter ETL-Service (Extrahieren, Transformieren und Laden), mit dem Sie Ihre Daten einfach und kostengünstig kategorisieren, bereinigen, anreichern und zuverlässig zwischen verschiedenen Datenspeichern verschieben können. Der AWS Glue Data Catalog bietet ein einheitliches Metadaten-Repository für eine Vielzahl von Datenquellen und Datenformaten, das sich in Amazon EMR sowie Amazon RDS, Amazon Redshift, Redshift Spectrum, Athena und jede mit dem Apache Hive Metastore kompatible Anwendung integrieren lässt. AWS Glue-Crawler können automatisch Schemas aus Quelldaten in Amazon S3 ableiten und die zugehörigen Metadaten im

Datenkatalog speichern. Weitere Informationen zum Datenkatalog finden Sie unter [Füllen des AWS Glue-Datenkatalogs im AWS Glue-Entwicklerhandbuch](#).

Für AWS Glue fallen separate Gebühren an. Es gibt eine monatliche Gebühr für das Speichern und Zugreifen auf die Metadaten im Datenkatalog, einen Stundensatz, der pro Minute für AWS Glue ETL-Jobs und Crawler-Laufzeit abgerechnet wird, und einen Stundensatz, der pro Minute für jeden bereitgestellten Entwicklungsendpunkt abgerechnet wird. Der Datenkatalog ermöglicht Ihnen die gebührenfreie Speicherung von bis zu einer Million Objekten. Wenn Sie mehr als eine Million Objekte speichern, fällt eine Gebühr von 1 USD für jeweils 100.000 zusätzliche Objekte an. Im Data Catalog ist ein Objekt eine Tabelle, eine Partition oder eine Datenbank. Weitere Informationen finden Sie unter [Glue-Preise](#).

#### Important

Wenn Sie vor dem 14. August 2017 Tabellen mit Amazon Athena oder Amazon Redshift Spectrum erstellt haben, werden Datenbanken und Tabellen in einem von Athena verwalteten Katalog gespeichert, der vom Glue-Datenkatalog getrennt ist. AWS Um Amazon EMR mit diesen Tabellen zu integrieren, müssen Sie ein Upgrade auf den AWS Glue Data Catalog durchführen. Weitere Informationen finden Sie unter [Upgrade auf den AWS Glue-Datenkatalog](#) im Amazon Athena Athena-Benutzerhandbuch.

## AWS Glue Data Catalog als Metastore angeben

Sie können den AWS Glue-Datenkatalog mithilfe der Amazon EMR-API, der AWS Management Console, der AWS CLI, oder als Metastore angeben. Wenn Sie die CLI oder API verwenden, verwenden Sie die Konfigurationsklassifizierung für Presto, um den Datenkatalog anzugeben. Darüber hinaus können Sie mit Amazon EMR 5.16.0 und höher die Konfigurationsklassifizierung verwenden, um einen Datenkatalog in einem anderen Format anzugeben. AWS-Konto Wenn Sie die Konsole verwenden, können Sie den Data Catalog mit den erweiterten Optionen oder den Schnelloptionen angeben.

### New console

Um AWS Glue Data Catalog mit der neuen Konsole als Hive-Metastore anzugeben

1. Melden Sie sich bei der AWS Management Console an und öffnen Sie die Amazon EMR-Konsole unter <https://console.aws.amazon.com/emr>.

2. Wählen Sie im linken Navigationsbereich unter EMR in EC2 die Option Cluster und dann Cluster erstellen aus.
3. Wählen Sie unter Anwendungspaket die Option Presto aus.
4. Aktivieren Sie unter Einstellungen für den AWS Glue Data Catalog das Kontrollkästchen Für Presto-Tabellenmetadaten verwenden.
5. Wählen Sie alle anderen Optionen aus, die für Ihren Cluster gelten.
6. Um Ihren Cluster jetzt zu starten, wählen Sie Cluster erstellen aus.

## Old console

Um den AWS Glue-Datenkatalog als Standard-Presto-Metastore mit der alten Konsole anzugeben

1. Navigieren Sie zur neuen Amazon-EMR-Konsole und wählen Sie in der Seitennavigation die Option Zur alten Konsole wechseln aus. Weitere Informationen darüber, was Sie erwartet, wenn Sie zur alten Konsole wechseln, finden Sie unter [Verwenden der alten Konsole](#).
2. Wählen Sie Create Cluster (Cluster erstellen) und Go to advanced options (Zu erweiterten Optionen) aus.
3. Wählen Sie unter Software Configuration eine Version von emr-5.10-0 oder höher und wählen Sie Presto.
4. Wählen Sie Use for Presto table metadata (Für Presto-Tabellen-Metadaten verwenden), wählen Sie Next (Weiter), und ergänzen Sie dann andere Einstellungen für Ihren Cluster, wie für Ihre Anwendung sinnvoll.

## CLI

Um den AWS Glue-Datenkatalog als Standard-Hive-Metastore anzugeben, verwenden Sie den AWS CLI

Beispiele dafür, wie Sie die folgenden Konfigurationsklassifizierungen unten angeben, wenn Sie einen Cluster erstellen, finden Sie unter [Anwendungen konfigurieren](#).

Amazon EMR 5.16.0 und höher

- Setzen Sie die Eigenschaft `hive.metastore` auf `glue` fest, wie in dem folgenden JSON-Beispiel gezeigt.

```
[
```



```
[
  {
    "Classification": "presto-connector-hive",
    "Properties": {
      "hive.metastore": "glue"
    }
  }
]
```

Um einen Datenkatalog in einem anderen Format anzugeben AWS-Konto, fügen Sie die `hive.metastore.glue.catalogid` Eigenschaft hinzu, wie im folgenden JSON-Beispiel gezeigt. Ersetzen Sie *acct-id* durch das AWS -Konto des Data Catalog. Die Verwendung eines Datenkatalog in einem anderen AWS-Konto ist für Amazon-EMR-Version 5.15.0 und früher nicht verfügbar.

```
[
  {
    "Classification": "presto-connector-hive",
    "Properties": {
      "hive.metastore": "glue",
      "hive.metastore.glue.catalogid": "acct-id"
    }
  }
]
```

#### Amazon EMR 5.10.0 bis 5.15.0

Setzen Sie die Eigenschaft `hive.metastore.glue.datacatalog.enabled` auf `true` fest, wie im folgenden JSON-Beispiel gezeigt.

```
[
  {
    "Classification": "presto-connector-hive",
    "Properties": {
      "hive.metastore.glue.datacatalog.enabled": "true"
    }
  }
]
```

#### Amazon EMR 6.1.0 und höher mit PrestoSQL (Trino)

Ab EMR Version 6.1.0 unterstützt PrestoSQL auch Glue als Standard-Hive-Metastore. Verwenden Sie die `prestoql-connector-hive`-Konfigurationsklassifizierung, um die Eigenschaft `hive.metastore` auf `glue` zu setzen, wie im folgenden JSON-Beispiel veranschaulicht.

Amazon-EMR-Versionen 6.4.0 und höher verwenden den neuen Namen Trino anstelle von PrestoSQL. Wenn Sie Trino verwenden, ersetzen Sie die Klassifizierung `prestoql-connector-hive` in der folgenden Konfiguration mit `trino-connector-hive` durch.

```
[
  {
    "Classification": "prestoql-connector-hive",
    "Properties": {
      "hive.metastore": "glue"
    }
  }
]
```

Um Metastores auf einem langlebigen Cluster zu wechseln, können Sie manuell die für die Version geeigneten Werte festlegen, indem Sie eine Verbindung zum Master-Knoten herstellen, die Eigenschaftswerte direkt in der Datei `/etc/presto/conf/catalog/hive.properties` bearbeiten und den Presto-Server (`sudo restart presto-server`) neu starten. Wenn Sie diese Methode bei Amazon-EMR-Version 5.15.0 oder früher verwenden, stellen Sie sicher, dass für `hive.table-statistics-enabled` der Wert `false` eingestellt ist. Diese Einstellung ist nicht erforderlich, wenn Sie Versionen ab 5.16.0 verwenden. Trotzdem werden Tabellen- und Partitionsstatistiken nicht unterstützt.

## IAM-Berechtigungen

Das EC2-Instanzprofil für einen Cluster muss über IAM-Berechtigungen für AWS Glue-Aktionen verfügen. Wenn Sie die Verschlüsselung für AWS Glue Data Catalog-Objekte aktivieren, muss die Rolle außerdem berechtigt sein, die für die Verschlüsselung AWS KMS key verwendeten Daten zu verschlüsseln, zu entschlüsseln und zu generieren.

## Berechtigungen für AWS Glue-Aktionen

Wenn Sie das standardmäßige EC2-Instance-Profil für Amazon EMR verwenden, ist keine Aktion erforderlich. Die `AmazonElasticMapReduceforEC2Role` verwaltete Richtlinie, die an die angehängt ist, `EMR_EC2_DefaultRole` ermöglicht alle erforderlichen AWS Glue-Aktionen. Wenn Sie jedoch ein benutzerdefiniertes EC2-Instanzprofil und Berechtigungen angeben, müssen Sie die entsprechenden AWS Glue-Aktionen konfigurieren. Verwenden Sie hierzu die verwaltete Richtlinie `AmazonElasticMapReduceforEC2Role` als Ausgangspunkt. Weitere Informationen finden Sie unter [Servicerolle für Cluster-EC2-Instances \(EC2-Instance-Profil\)](#) im Verwaltungshandbuch für Amazon EMR.

## Berechtigungen zum Verschlüsseln und Entschlüsseln von AWS Glue Data Catalog

Ihr Instance-Profil benötigt die Erlaubnis, Daten mithilfe Ihres Schlüssels zu verschlüsseln und zu entschlüsseln. Sie müssen diese Berechtigungen nicht konfigurieren, wenn die beiden folgenden Aussagen zutreffen:

- Sie aktivieren die Verschlüsselung für AWS Glue Data Catalog-Objekte mithilfe von verwalteten Schlüsseln für AWS Glue.
- Sie verwenden einen Cluster, der sich im selben Verzeichnis befindet AWS-Konto wie der AWS Glue-Datenkatalog.

Andernfalls müssen Sie die folgende Anweisung der Berechtigungsrichtlinie hinzufügen, die mit Ihrem EC2-Instance-Profil verknüpft ist.

```
[
  {
    "Version": "2012-10-17",
    "Statement": [
      {
        "Effect": "Allow",
        "Action": [
          "kms:Decrypt",
          "kms:Encrypt",
          "kms:GenerateDataKey"
        ],
        "Resource": "arn:aws:kms:region:acct-
id:key/12345678-1234-1234-1234-123456789012"
      }
    ]
  }
]
```

```
}  
]
```

Weitere Informationen zur Verschlüsselung von AWS Glue Data Catalog finden Sie unter [Verschlüsselung Ihres Datenkatalogs](#) im AWS Glue Developer Guide.

## Ressourcenbasierte Berechtigungen

Wenn Sie AWS Glue in Verbindung mit Hive, Spark oder Presto in Amazon EMR verwenden, unterstützt AWS Glue ressourcenbasierte Richtlinien zur Steuerung des Zugriffs auf Datenkatalogressourcen. Zu diesen Ressourcen gehören Datenbanken, Tabellen, Verbindungen und benutzerdefinierte Funktionen. Weitere Informationen finden Sie unter [Verwenden von ressourcenbasierten Richtlinien für AWS Glue](#) im AWS -Glue-Entwicklerhandbuch.

Wenn Sie ressourcenbasierte Richtlinien verwenden, um den Zugriff auf AWS Glue von Amazon EMR aus zu beschränken, muss der Principal, den Sie in der Berechtigungsrichtlinie angeben, der Rollen-ARN sein, der dem EC2-Instance-Profil zugeordnet ist, das bei der Erstellung eines Clusters angegeben wird. Beispielsweise können Sie für eine ressourcenbasierte Richtlinie, die an einen Katalog angehängt ist, den Rollen-ARN für die Standarddienstrolle für Cluster-EC2-Instances, *EMR\_EC2\_*, DefaultRole als angeben, indem Sie das im folgenden Beispiel gezeigte Principal Format verwenden:

```
arn:aws:iam::acct-id:role/EMR_EC2_DefaultRole
```

Die *Konto-ID* kann sich von der AWS Glue-Konto-ID unterscheiden. Dies ermöglicht den Zugriff von EMR-Clustern in verschiedenen Konten aus. Sie können mehrere Principals angeben, von denen jeder aus einem anderen Konto stammt.

## Überlegungen zur Verwendung von AWS Glue Data Catalog

Beachten Sie die folgenden Punkte, wenn Sie AWS Glue Data Catalog als Metastore mit Presto verwenden:

- Das Umbenennen von Tabellen in AWS Glue wird nicht unterstützt.
- Wenn Sie eine Hive-Tabelle erstellen, ohne eine LOCATION anzugeben, werden die Tabellendaten an dem in der Eigenschaft `hive.metastore.warehouse.dir` angegebenen Ort gespeichert. Standardmäßig ist dies ein Speicherort in HDFS. Wenn ein anderer Cluster auf die Tabelle zugreifen muss, schlägt er fehl, sofern er nicht über ausreichende Berechtigungen für den Cluster

verfügt, der die Tabelle erstellt hat. Da HDFS-Speicher vorübergehend ist, gehen außerdem die Tabellendaten verloren, wenn der Cluster beendet wird, und die Tabelle muss neu erstellt werden. Wir empfehlen, dass Sie LOCATION in Amazon S3 angeben, wenn Sie eine Hive-Tabelle mit AWS Glue erstellen. Alternativ können Sie die hive-site-Konfigurationsklassifizierung verwenden, um einen Speicherort in Amazon S3 für hive.metastore.warehouse.dir anzugeben, der für alle Hive-Tabellen gilt. Wenn eine Tabelle an einem HDFS-Speicherort erstellt wurde und der Cluster, der sie erstellt hat, noch läuft, können Sie den Tabellenspeicherort von AWS Glue aus auf Amazon S3 aktualisieren. Weitere Informationen finden Sie unter [Arbeiten mit Tabellen auf der AWS Glue-Konsole](#) im AWS Glue-Entwicklerhandbuch.

- Partitionswerte, die Anführungszeichen und Apostrophe enthalten, werden nicht unterstützt, zum Beispiel PARTITION (owner="Doe's").
- [Spaltenstatistiken](#) werden für emr-5.31.0 und höher unterstützt.
- Die Verwendung der [Hive-Autorisierung](#) wird nicht unterstützt. Als Alternative sollten Sie die Verwendung [ressourcenbasierter Richtlinien für AWS Glue](#) in Betracht ziehen. Weitere Informationen finden Sie unter [Verwenden von ressourcenbasierten Richtlinien für Amazon EMR Access to AWS Glue](#) Data Catalog.

## S3-Select-Pushdown mit Presto zur Verbesserung der Leistung verwenden

Mit Amazon EMR Version 5.18.0 und höher können Sie [S3-Select](#)-Pushdown mit Presto in Amazon EMR verwenden. Dieses Feature ermöglicht Presto Rechenarbeit der Projektionsoperationen (z. B. SELECT) und der Prädikationoperationen (z. B. WHERE) in Amazon S3 herunter zu schieben („Pushdown“). Auf diese Weise können Abfragen die aus Amazon S3 abgerufenen Daten auf die benötigten Daten beschränken, was die Leistung verbessern und die Datenmenge reduzieren kann, die in bestimmten Anwendungen zwischen Amazon EMR und Amazon S3 übertragen wird.

### Ist S3-Select-Pushdown das Richtige für meine Anwendung?

Wir empfehlen, dass Sie Benchmark-Tests für Ihre Anwendungen im Vergleich mit und ohne S3 Select-Pushdown durchführen, um zu sehen, ob es für Ihre Anwendung geeignet sein könnte.

Verwenden Sie die folgenden Richtlinien, um zu bestimmen, ob Ihre Anwendung ein Kandidat für die Verwendung von S3 ist:

- Ihre Abfrage filtert mehr als die Hälfte des ursprünglichen Datensatzes.

- Ihre Abfrage-Filterprädikate verwenden Spalten mit einem Datentyp, der von Presto und S3 Select unterstützt wird. Die Datentypen `timestamp`, `real` und `double` werden von S3 Select-Pushdown nicht unterstützt. Wir empfehlen für numerische Daten die Verwendung des Datentyps `decimal`. Weitere Informationen zu unterstützten Datentypen für S3 Select finden Sie unter [Datentypen](#) im Benutzerhandbuch für Amazon Simple Storage Service.
- Ihre Netzwerkverbindung zwischen Amazon S3 und dem Amazon-EMR-Cluster hat eine gute Übertragungsgeschwindigkeit und verfügbare Bandbreite. Amazon S3 komprimiert keine HTTP-Antworten, sodass die Antwortgröße bei komprimierten Eingabedateien wahrscheinlich zunimmt.

## Überlegungen und Einschränkungen

- Es werden nur im CSV-Format gespeicherte Objekte unterstützt. Objekte können optional mit `gzip` oder `bzip2` komprimiert bzw. dekomprimiert werden.
- Die Eigenschaft `AllowQuotedRecordDelimiters` wird nicht unterstützt. Wenn diese Eigenschaft angegeben ist, schlägt die Abfrage fehl.
- Die Amazon-S3-serverseitige Verschlüsselung mit vom Kunden bereitgestellten Verschlüsselungsschlüsseln (SSE-C) und die clientseitige Verschlüsselung werden nicht unterstützt.
- S3 Select-Pushdown ist kein Ersatz für die Verwendung von spaltenbasierten oder komprimierten Dateiformaten wie z. B. ORC oder Parquet.

## Aktivieren von S3-Select-Pushdown für Presto oder Trino

Um S3-Select-Pushdown für Presto auf Amazon EMR zu aktivieren, verwenden Sie die `presto-connector-hive`-Konfigurationsklassifizierung, um `hive.s3select-pushdown.enabled` auf `true` zu setzen, wie im nachfolgenden Beispiel gezeigt. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#). Außerdem muss der Wert für „`hive.s3select-pushdown.max-connections`“ festgelegt werden. Für die meisten Anwendungen sollte die Standardeinstellung **500** ausreichen. Weitere Informationen hierzu finden Sie unter [Verstehen und Optimieren von „hive.s3select-pushdown.max-connections“](#) weiter unten.

Ersetzen Sie für PrestoSQL auf den EMR-Versionen 6.1.0 bis 6.3.0 im folgenden Beispiel `presto-connector-hive` durch `prestoql-connector-hive`.

Amazon-EMR-Versionen 6.4.0 und höher verwenden den neuen Namen Trino anstelle von PrestoSQL. Wenn Sie Trino verwenden, ersetzen Sie im Beispiel unten *presto-connector-hive* durch `trino-connector-hive`

```
[
  {
    "classification": "presto-connector-hive",
    "properties": {
      "hive.s3select-pushdown.enabled": "true",
      "hive.s3select-pushdown.max-connections": "500"
    }
  }
]
```

## Verstehen und Optimieren von „hive.s3select-pushdown.max-connections“

Standardmäßig verwendet Presto EMRFS als Dateisystem. Die Einstellung `fs.s3.maxConnections` in der `emrfs-site`-Konfigurationsklassifizierung gibt die maximal zulässige Anzahl an Clientverbindungen zu Amazon S3 über EMRFS für Presto an. Standardmäßig ist dies 500. S3-Select-Pushdown umgeht beim Zugriff auf Amazon S3 für Prädikatoperationen EMRFS. In diesem Fall gibt der Wert von `hive.s3select-pushdown.max-connections` die maximale Anzahl von Clientverbindungen an, die für diese Operationen von Worker-Knoten aus zulässig ist. Für alle Anfragen an Amazon S3, die Presto initiiert und die nicht nach unten verschoben werden (z. B. GET-Operationen), gilt jedoch weiterhin der Wert von `fs.s3.maxConnections`.

Wenn Ihre Anwendung den Fehler „Timeout waiting for connection from pool (Timeout beim Warten auf eine Verbindung aus dem Pool)“ erhält, erhöhen Sie den Wert von `hive.s3select-pushdown.max-connections` und von `fs.s3.maxConnections`.

## Hinzufügen von Datenbank-Konnektoren

Sie können JDBC-Konnektoreigenschaften mithilfe von Konfigurationsklassifizierungen konfigurieren, wenn Sie einen Cluster erstellen. Konfigurationsklassifizierungen beginnen mit `presto-connector`, z. B. `presto-connector-postgresql`. Die verfügbaren Konfigurationsklassifizierungen variieren je nach Amazon-EMR-Version. Die mit der neuesten Release-Version verfügbaren Konfigurationsklassifizierungen finden Sie unter [the section called “5.36.2 Konfigurationsklassifizierungen”](#) Amazon EMR 5.36.2. Wenn Sie eine andere Version von Amazon EMR verwenden, finden Sie Informationen zu den Konfigurationsklassifizierungen unter [Amazon-EMR-5.x-Versionen](#). Weitere Informationen zu den Eigenschaften, die mit einzelnen

Konnektoren konfiguriert werden können, finden Sie unter <https://prestodb.io/docs/current/connector.html>.

### Example – Konfigurieren eines Clusters mit dem PostgreSQL-JDBC-Konnektor

Zum Starten eines Clusters mit installiertem und konfiguriertem PostgreSQL-Konnektor erstellen Sie zunächst eine JSON-Datei mit der Konfigurationsklassifizierung – beispielsweise `myConfig.json` – mit dem folgenden Inhalt und speichern Sie sie lokal.

Ersetzen Sie die Verbindungseigenschaften entsprechend Ihrer Umgebung und wie im Thema [PostgreSQL-Konnektor](#) in der Presto-Dokumentation beschrieben.

```
[
  {
    "Classification": "presto-connector-postgresql",
    "Properties": {
      "connection-url": "jdbc:postgresql://example.net:5432/database",
      "connection-user": "MYUSER",
      "connection-password": "MYPASS"
    },
    "Configurations": []
  }
]
```

Wenn Sie den Cluster erstellen, geben Sie den Pfad zu der JSON-Datei unter Verwendung der `--configurations`-Option an, wie im folgenden Beispiel gezeigt, wobei `myConfig.json` in demselben Verzeichnis liegt, in dem Sie den Befehl ausführen:

```
aws emr create-cluster --name PrestoConnector --release-label emr-5.36.2 --instance-
type m5.xlarge \
--instance-count 2 --applications Name=Hadoop Name=Hive Name=Pig Name=Presto \
--use-default-roles --ec2-attributes KeyName=myKey \
--log-uri s3://my-bucket/logs --enable-debugging \
--configurations file://myConfig.json
```

## Verwenden von SSL/TLS und Konfigurieren von LDAPS mit Presto in Amazon EMR

Mit Amazon EMR ab Version 5.6.0 können Sie zum Schutz der internen Kommunikation zwischen Presto-Knoten SSL/TLS [gesicherte interne Kommunikation](#) aktivieren. Zu diesem Zweck richten



Sie eine Sicherheitskonfiguration für die Verschlüsselung der Daten während der Übertragung ein. Weitere Informationen finden Sie unter [Verschlüsselungsoptionen](#) und [Sicherheitskonfigurationen zur Einrichtung der Clustersicherheit verbinden](#) im Verwaltungshandbuch für Amazon EMR.

Wenn Sie eine Sicherheitskonfiguration mit Verschlüsselung der Daten während der Übertragung verwenden, führt Amazon EMR Folgendes für Presto durch:

- Der Service verteilt die von Ihnen angegebenen Verschlüsselungsartefakte oder Zertifikate über den Presto-Cluster. Weitere Informationen finden Sie unter [Bereitstellen von Zertifikaten für die Datenverschlüsselung während der Übertragung](#).
- Legt die folgenden Eigenschaften mit der Konfigurationsklassifizierung `presto-config` fest, die der `config.properties`-Datei für Presto entspricht:
  - Setzt `http-server.http.enabled` auf `false` auf allen Knoten, wodurch HTTP zugunsten von HTTPS deaktiviert wird. Dazu müssen Sie Zertifikate angeben, die für öffentliches und privates DNS funktionieren, wenn Sie die Sicherheitskonfiguration für die Verschlüsselung während der Übertragung einrichten. Eine Möglichkeit, dies zu tun, besteht darin, SAN-Zertifikate (Subject Alternative Name) zu verwenden, die mehrere Domains unterstützen.
  - Legt `http-server.https.*`-Werte fest. Konfigurationsdetails finden Sie unter [LDAP-Authentifizierung](#) in der Presto-Dokumentation.
- Für PrestoSQL (Trino) auf EMR-Version 6.1.0 und höher konfiguriert Amazon EMR automatisch einen gemeinsamen geheimen Schlüssel für die sichere interne Kommunikation zwischen Clusterknoten. Sie müssen keine zusätzliche Konfiguration vornehmen, um dieses Sicherheitsfeature zu aktivieren, und Sie können die Konfiguration mit Ihrem eigenen geheimen Schlüssel überschreiben. Informationen zur internen Authentifizierung von Trino finden Sie in der [Trino-353-Dokumentation: Sichere interne Kommunikation](#).

Außerdem können Sie bei Amazon-EMR-Version 5.10.0 und höher die [LDAP-Authentifizierung](#) für Clientverbindungen mit Presto-Koordinator über HTTPS einrichten. Bei dieser Einrichtung wird sicheres LDAP (LDAPS) verwendet. TLS muss auf dem LDAP-Server aktiviert werden und der Presto-Cluster muss eine Sicherheitskonfiguration mit aktivierter Verschlüsselung der Daten bei der Übertragung verwenden. Zusätzliche Konfigurationsschritte sind erforderlich. Die Konfigurationsoptionen unterscheiden sich in Abhängigkeit von der verwendeten Amazon-EMR-Version. Weitere Informationen finden Sie unter [LDAP-Authentifizierung für Presto in Amazon EMR verwenden](#).

Presto in Amazon EMR verwendet für internes HTTPS standardmäßig Port 8446. Der für die interne Kommunikation verwendete Port muss mit dem Port für den Client-Zugriff über HTTPS

auf den Presto-Koordinator identisch sein. Die Eigenschaft `http-server.https.port` in der Konfigurationsklassifizierung `presto-config` gibt den Port an.

## LDAP-Authentifizierung für Presto in Amazon EMR verwenden

Befolgen Sie zum Konfigurieren von LDAP die Schritte in diesem Abschnitt. Beachten Sie die einzelnen Schritte sowie die Beispiele und Links mit weiterführenden Informationen.

Schritte zum Konfigurieren der LDAP-Authentifizierung

- [Schritt 1: Informationen zum LDAP-Server erfassen und Serverzertifikat zu Amazon S3 kopieren](#)
- [Schritt 2: Sicherheitskonfiguration einrichten](#)
- [Schritt 3: JSON-Konfigurationsobjekt mit Presto-Eigenschaften für LDAP erstellen](#)
- [Schritt 4: Skript zum Kopieren des LDAP-Serverzertifikats erstellen und nach Amazon S3 hochladen](#)
- [Schritt 5: Cluster erstellen](#)

### Schritt 1: Informationen zum LDAP-Server erfassen und Serverzertifikat zu Amazon S3 kopieren

Sie benötigen die im folgenden Abschnitt angegebenen Informationen und Elemente vom LDAP-Server zum Konfigurieren der LDAP-Authentifizierung.

Die IP-Adresse oder der Host-Name des LDAP-Servers

Der Presto-Koordinator auf dem Amazon-EMR-Hauptknoten muss den LDAP-Server an der angegebenen IP-Adresse oder unter dem angegebenen Host-Namen erreichen können.

Standardmäßig kommuniziert Presto mit dem LDAP-Server unter Verwendung von LDAPS über Port 636. Wenn Ihre LDAP-Implementierung einen benutzerdefinierten Port benötigt, können Sie diesen bei Verwendung von Amazon EMR ab Version 5.16.0 mit der Eigenschaft `ldap.url` angeben.

Verwenden Sie `authentication.ldap.url` bei früheren Versionen. Ersetzen Sie 636 durch den benutzerdefinierten Port, wie in den `presto-config`-Konfigurationsklassifizierungsbeispielen unter [Schritt 3: JSON-Konfigurationsobjekt mit Presto-Eigenschaften für LDAP erstellen](#) gezeigt. Stellen Sie sicher, dass alle Firewalls und Sicherheitsgruppen ein- und ausgehenden Datenverkehr über Port 636 (oder den benutzerdefinierten Port) und zudem Port 8446 (oder den benutzerdefinierten Port), der für die interne Clusterkommunikation verwendet wird, zulassen.

## Das LDAP-Serverzertifikat

Sie müssen die Zertifikatsdatei in einen sicheren Speicherort in Amazon S3 hochladen. Weitere Informationen finden Sie unter [Wie lade ich Dateien und Ordner in einen S3 Bucket hoch?](#) im Benutzerhandbuch für Amazon Simple Storage Service. Sie erstellen eine Bootstrap-Aktion, die dieses Zertifikat aus Amazon S3 auf jeden Knoten im Cluster kopiert, wenn der Cluster gestartet wird. Wählen Sie unter [Schritt 4: Skript zum Kopieren des LDAP-Serverzertifikats erstellen und nach Amazon S3 hochladen](#). Das Beispielzertifikat ist `s3:///ldap_server.crt`. MyBucket

## Die LDAP-Servereinstellungen für anonyme Bindung

Wenn die anonyme Bindung auf PrestoDB deaktiviert ist, benötigen Sie die Benutzer-ID (UID) und das Passwort eines Kontos mit Berechtigungen, die Bindung an den LDAP-Server herzustellen, damit der PrestoDB-Server eine Verbindung herstellen kann. Sie geben die UID und das Passwort mit den Eigenschaften `internal-communication.ldap.user` und `internal-communication.ldap.password` in der Konfigurationsklassifizierung `presto-config` an. Amazon EMR 5.10.0 unterstützt diese Einstellungen nicht, sodass die anonyme Bindung auf dem LDAP-Server unterstützt werden muss, wenn Sie diese Version verwenden.

Beachten Sie, dass Trino die Konfiguration der anonymen Bindung nicht benötigt.

So rufen Sie den Status der anonymen Bindung auf dem LDAP-Server ab

- Verwenden Sie von einem Linux-Client aus den Befehl [ldapwhoami](#), wie im folgenden Beispiel gezeigt:

```
ldapwhoami -x -H ldaps://LDAPServerHostNameOrIPAddress
```

Wenn die anonyme Bindung nicht zulässig ist, gibt der Befehl Folgendes zurück:

```
ldap_bind: Inappropriate authentication (48)  
additional info: anonymous bind disallowed
```

So verifizieren Sie, dass ein Konto über die Berechtigungen für einen LDAP-Server verfügt, der die einfache Authentifizierung verwendet

- Verwenden Sie von einem Linux-Client aus den Befehl [ldapwhoami](#), wie im folgenden Beispiel gezeigt. *Das Beispiel verwendet einen fiktiven Benutzer, Presto, der auf einem Open LDAP-Server gespeichert ist, der auf einer EC2-Instance*

mit dem fiktiven Hostnamen `-xxx.ec2.internal` läuft. `ip-xxx-xxx-xxx` Dem Benutzer ist die Organisationseinheit (OU) `admins` und das Passwort `123456` zugeordnet:

```
ldapwhoami -x -w "123456" -D uid=presto,ou=admins,dc=ec2,dc=internal -H ldaps://ip-xxx-xxx-xxx-xxx.ec2.internal
```

Wenn das Konto gültig ist und über die erforderlichen Berechtigungen verfügt, gibt der Befehl Folgendes zurück:

```
dn:uid=presto,ou=admins,dc=ec2,dc=internal
```

Die Beispielkonfigurationen unter [Schritt 3: JSON-Konfigurationsobjekt mit Presto-Eigenschaften für LDAP erstellen](#) enthalten dieses Konto, um die Verständlichkeit zu verbessern. Ausgenommen ist das 5.10.0-Beispiel, weil es hier nicht unterstützt wird. Wenn der LDAP-Server die anonyme Bindung verwendet, entfernen Sie die Name-Wert-Paare `internal-communication.ldap.user` und `internal-communication.ldap.password`.

Der eindeutige LDAP-Name (DN) für Presto-Benutzer

Wenn Sie die LDAP-Konfiguration für Presto angeben, geben Sie ein Bindungsmuster an, das aus `${USER}` und einer Organisationseinheit (OU) sowie zusätzlichen Domainkomponenten (DCs) besteht. Presto ersetzt `${USER}` im Rahmen der Passwort-Authentifizierung durch die tatsächliche Benutzer-ID (UID) jedes Benutzers, damit der Wert dem eindeutigen Namen (DN) entspricht, den dieses Bindungsmuster angibt. Sie benötigen die OUs, zu denen qualifizierte Benutzer gehören, sowie deren DCs. Um Benutzern aus der `admins`-OU in der `corp.example.com`-Domain zu gestatten, sich bei Presto zu authentifizieren, geben Sie `${USER},ou=admins,dc=corp,dc=example,dc=com` als Benutzerbindungsmuster an.

#### Note

Wenn Sie verwenden AWS CloudFormation, müssen Sie die Funktion `Fn::Sub` verwenden, um sie durch die tatsächliche Benutzer-ID (UID) zu ersetzen. `${USER}` Weitere Informationen finden Sie im Thema [Fn::Sub](#) im AWS CloudFormation -Benutzerhandbuch.

Bei Verwendung von Amazon EMR 5.10.0, können Sie nur ein solches Muster angeben. Bei Verwendung von Amazon EMR ab 5.11.0 oder höher können Sie mehrere Muster angeben, die durch einen Doppelpunkt (:) voneinander getrennt werden. Benutzer, die sich bei Presto zu authentifizieren versuchen, werden mit den ersten Muster verglichen, dann mit dem zweiten usw. Ein Beispiel finden Sie unter [Schritt 3: JSON-Konfigurationsobjekt mit Presto-Eigenschaften für LDAP erstellen](#).

## Schritt 2: Sicherheitskonfiguration einrichten

Erstellen Sie eine Sicherheitskonfiguration mit aktivierter Datenverschlüsselung während der Übertragung. Weitere Informationen finden Sie unter [Erstellen einer Sicherheitskonfiguration](#) im Verwaltungshandbuch für Amazon EMR. Die Verschlüsselungsartefakte, die Sie bereitstellen, wenn Sie die Verschlüsselung der Daten während der Übertragung einrichten, werden zum Verschlüsseln der internen Kommunikation zwischen Presto-Knoten verwendet. Weitere Informationen finden Sie unter [Bereitstellen von Zertifikaten für die Datenverschlüsselung während der Übertragung](#). Das LDAP-Serverzertifikat wird verwendet, um Client-Verbindungen zum Presto-Server zu authentifizieren.

## Schritt 3: JSON-Konfigurationsobjekt mit Presto-Eigenschaften für LDAP erstellen

Sie verwenden die Konfigurationsklassifizierung `presto-config` zum Festlegen von Presto-Eigenschaften für LDAP. Das Format und der Inhalt von `presto-config` unterscheiden sich je nach der Amazon-EMR-Release-Version und der Presto-Installation (PrestoDB oder Trino) geringfügig. Beispiele für Konfigurationsunterschiede finden Sie unten in diesem Abschnitt. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

*Bei den folgenden Schritten wird davon ausgegangen, dass Sie die JSON-Daten in einer Datei, `.json`, `MyPrestoConfig` speichern.* Wenn Sie die Konsole verwenden, laden Sie die Datei in einen sicheren Speicherort in Amazon S3 hoch, damit Sie beim Erstellen des Clusters darauf verweisen können. Wenn Sie die verwenden AWS CLI, können Sie lokal auf die Datei verweisen.

Example Amazon EMR 6.1.0 und höher mit PrestoSQL (Trino)

Das folgende Beispiel verwendet die LDAP und das Host-Namen aus [Schritt 1: Informationen zum LDAP-Server erfassen und Serverzertifikat zu Amazon S3 kopieren](#), um den LDAP-Server für die Bindung zu authentifizieren. Zwei Benutzerbindungsmuster werden angegeben. Dies bedeutet, dass Benutzer in der OU `admins` und der OU `datascientists` auf dem LDAP-Server für die Authentifizierung als Benutzer gegenüber dem Trino-Server qualifiziert sind. Die Bindungsmuster werden durch einen Doppelpunkt (:) voneinander getrennt.

Amazon-EMR-Versionen 6.4.0 und höher verwenden den neuen Namen Trino anstelle von PrestoSQL. Wenn Sie Trino verwenden, ersetzen Sie die Klassifizierung *prestoql-config* in der folgenden Konfiguration durch *trino-config* und *prestoql-password-authenticator* durch *trino-password-authenticator*.

```
[
  {
    "Classification": "prestoql-config",
    "Properties": {
      "http-server.authentication.type": "PASSWORD"
    }
  },
  {
    "Classification": "prestoql-password-authenticator",
    "Properties": {
      "password-authenticator.name": "ldap",
      "ldap.url": "ldaps://ip-xxx-xxx-xxx-xxx.ec2.internal:636",
      "ldap.user-bind-pattern": "uid=${USER},ou=admins,dc=ec2,dc=internal:uid=${USER},ou=datascientists,dc=ec2,dc=internal"
    }
  }
]
```

### Example Amazon EMR 5.16.0 und höher

Das folgende Beispiel verwendet die LDAP-Benutzer-ID und das Passwort sowie den LDAP-Host-Namen aus [Schritt 1: Informationen zum LDAP-Server erfassen und Serverzertifikat zu Amazon S3 kopieren](#), um den LDAP-Server für die Bindung zu authentifizieren. Zwei Benutzerbindungsmuster werden angegeben. Dies bedeutet, dass Benutzer in der OU admins und der OU datascientists auf dem LDAP-Server für die Authentifizierung als Benutzer gegenüber dem Presto-Server qualifiziert sind. Die Bindungsmuster werden durch einen Doppelpunkt (:) voneinander getrennt.

```
[{
  "Classification": "presto-config",
  "Properties": {
    "http-server.authentication.type": "PASSWORD"
  }
},
{
  "Classification": "presto-password-authenticator",
  "Properties": {
    "password-authenticator.name": "ldap",
```

```

        "ldap.url": "ldaps://ip-xxx-xxx-xxx-xxx.ec2.internal:636",
        "ldap.user-bind-pattern": "uid=
${USER},ou=admins,dc=ec2,dc=internal:uid=${USER},ou=datascientists,dc=ec2,dc=internal",
        "internal-communication.authentication.ldap.user": "presto",
        "internal-communication.authentication.ldap.password": "123456"
    }
}]]

```

### Example Amazon EMR 5.11.0 bis 5.15.0

Das Format der Konfigurationsklassifizierung `presto-config` unterscheidet sich geringfügig bei diesen Versionen. Das folgende Beispiel gibt dieselben Parameter an, die auch im vorherigen Beispiel verwendet wurden.

```

[ {
  "Classification": "presto-config",
  "Properties": {
    "http-server.authentication.type": "LDAP",
    "authentication.ldap.url": "ldaps://ip-xxx-xxx-xxx-
xxx.ec2.internal:636",
    "authentication.ldap.user-bind-pattern": "uid=
${USER},ou=admins,dc=ec2,dc=internal:uid=${USER},ou=datascientists,dc=ec2,dc=internal",
    "internal-communication.authentication.ldap.user": "presto",
    "internal-communication.authentication.ldap.password": "123456"
  }
}]]

```

### Example Amazon EMR 5.10.0

Amazon EMR 5.10.0 unterstützt nur die anonyme Bindung, diese Einträge werden deshalb nicht angegeben. Darüber hinaus kann nur ein Bindungsmuster angegeben werden.

```

[ {
  "Classification": "presto-config",
  "Properties": {
    "http-server.authentication.type": "LDAP",
    "authentication.ldap.url": "ldaps://ip-xxx-xxx-xxx-
xxx.ec2.internal:636",
    "ldap.user-bind-pattern": "uid=
${USER},ou=prestousers,dc=ec2,dc=internal"
  }
}]]

```

## Schritt 4: Skript zum Kopieren des LDAP-Serverzertifikats erstellen und nach Amazon S3 hochladen

Erstellen Sie ein Skript, das die Zertifikatdatei auf jeden Knoten im Cluster kopiert und in den Schlüsselspeicher einfügt. Erstellen Sie das Skript mit einem Texteditor, speichern Sie es und laden Sie es dann zu Amazon S3 hoch. In [Schritt 5: Cluster erstellen](#) wird auf die Skriptdatei als `s3://MyBucket/LoadLDAPCert.sh` verwiesen.

Im folgenden Beispielskript wird das Passwort `changeit` für den Standard-Schlüsselspeicher verwendet. Wir empfehlen, dass Sie eine Verbindung zum Master-Knoten herstellen, nachdem Sie den Cluster erstellt und das Passwort für den Schlüsselspeicher mit dem Befehl `keytool` geändert haben.

```
#!/bin/bash
aws s3 cp s3://MyBucket/ldap_server.crt .
sudo keytool -import -keystore /usr/lib/jvm/jre-1.8.0-openjdk.x86_64/lib/security/cacerts -trustcacerts -alias ldap_server -file ./ldap_server.crt -storepass changeit -noprompt
```

## Schritt 5: Cluster erstellen

Wenn Sie den Cluster erstellen, geben Sie Presto und andere Anwendungen an, die Amazon EMR installieren soll. Die folgenden Beispiele verweisen auch auf die Konfigurationsklassifizierungseigenschaften in einem JSON-Objekt. Sie können die Konfigurationsklassifizierung aber auch im Code angeben.

So erstellen Sie einen Presto-Cluster mit LDAP-Authentifizierung unter Verwendung der Amazon-EMR-Konsole


1. Navigieren Sie zur neuen Amazon-EMR-Konsole und wählen Sie in der Seitennavigation die Option [Zur alten Konsole wechseln](#) aus. Weitere Informationen darüber, was Sie erwartet, wenn Sie zur alten Konsole wechseln, finden Sie unter [Verwenden der alten Konsole](#).
2. Wählen Sie [Create Cluster \(Cluster erstellen\)](#) und [Go to advanced options \(Zu erweiterten Optionen\)](#) aus.
3. Wählen Sie Presto und andere Anwendungen, die Amazon EMR installieren soll. Wählen Sie unter [Software-Konfiguration](#) die zu verwendende Version von Amazon EMR. Die LDAP-Authentifizierung wird erst ab Amazon EMR 5.10.0 unterstützt.



4. Wählen Sie unter Software-Einstellungen bearbeiten die Option JSON aus S3 laden, geben Sie den Speicherort der in [Schritt 3: JSON-Konfigurationsobjekt mit Presto-Eigenschaften für LDAP erstellen](#) erstellten JSON-Konfigurationsdatei in Amazon S3 ein und wählen Sie dann Weiter.
5. Konfigurieren Sie die Cluster-Hardware und die Netzwerkfunktionen. Wählen Sie dann Next (Weiter).
6. Wählen Sie Bootstrap Actions (Bootstrap-Aktionen). Wählen Sie für Add bootstrap action (Bootstrap-Aktion hinzufügen) die Option Custom action (Benutzerdefinierte Aktion) und wählen Sie dann Configure and add (Konfigurieren und hinzufügen).
7. Geben Sie einen Namen für die Bootstrap-Aktion und den Speicherort für das Skript ein, in dem Sie das Skript erstellt haben [Schritt 4: Skript zum Kopieren des LDAP-Serverzertifikats erstellen und nach Amazon S3 hochladen](#), z. B. `s3://MyBucket/LoadLDAPCert.sh`, und wählen Sie dann Hinzufügen aus.
8. Wählen Sie unter General Options (Allgemeine Optionen), Tags und Additional Options (Weitere Optionen) für die Anwendung geeignete Einstellungen. Klicken Sie dann auf Next (Weiter).
9. Wählen Sie Authentication and encryption (Authentifizierung und Verschlüsselung) und dann die Security configuration (Sicherheitskonfiguration), die Sie in [Schritt 2: Sicherheitskonfiguration einrichten](#) erstellt haben.
10. Wählen Sie weitere für Ihre Anwendung geeignete Sicherheitsoptionen und wählen Sie dann Create cluster (Cluster erstellen).

So erstellen Sie einen Presto-Cluster mit LDAP-Authentifizierung unter Verwendung der AWS CLI

- Verwenden Sie den `aws emr create-cluster`-Befehl. Geben Sie mindestens die Presto-Anwendung und auch die Presto-Konfigurationsklassifizierung, das Bootstrap-Skript und die Sicherheitskonfiguration an, die Sie in den vorherigen Schritten erstellt haben. Das folgende Beispiel verweist auf die als JSON-Datei gespeicherte Konfigurationsdatei in dem Verzeichnis, in dem der Befehl ausgeführt wird. Das Bootstrap-Skript muss dagegen in Amazon S3 gespeichert werden. Im folgenden Beispiel wird verwendet `s3://MyBucket/LoadLDAPCert.sh`.

 Note

Linux-Zeilenfortsetzungszeichen (`\`) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (`^`).

```
aws emr create-cluster --applications Name=presto --release-label emr-5.16.0 \
--use-default-roles --ec2-attributes KeyName=MyKeyPair,SubnetId=subnet-1234ab5
\ --instance-count 3 --instance-type m5.xlarge --region us-west-2 --name
"MyPrestoWithLDAPAuth" \
--bootstrap-actions Name="Distribute LDAP server cert",Path="s3://MyBucket/
LoadLDAPCert.sh" \
--security-configuration MyPrestoLDAPSecCfg --configurations file://
MyPrestoConfig.json
```

## Den strikten Presto-Modus aktivieren

In bestimmten Situationen können Abfragen mit langer Laufzeit zu hohen Kosten führen und dazu führen, dass Amazon EMR mehr Cluster-Ressourcen verwendet. Dadurch werden anderen Workloads im Cluster Ressourcen entzogen. In den Versionen 6.8 und höher von Amazon EMR können Sie ein Feature im strikten Modus verwenden, das die folgenden Arten von Abfragen mit langer Laufzeit ablehnt oder Sie davor warnt:

- Abfragen ohne Prädikate für die partitionierten Spalten, die zu Tabellenscans großer Datenmengen führen
- Abfragen mit Kreuzverknüpfungen zwischen zwei großen Tabellen
- Abfragen, die eine große Anzahl von Zeilen ohne Limit sortieren

Nachdem Presto den Abfrageplan vollständig optimiert hat, wird der strikte Modus ausgeführt. Um den strikten Modus zu verwenden und an Ihre Abfrageanforderungen anzupassen, können Sie Presto wie folgt konfigurieren.

### Presto-Konfigurationen für den strikten Modus

Einstellung	Beschreibung	Standard
<code>strict-mode-enabled</code>	Schaltet den strikten Modus ein und aus. Der Wert von <code>true</code> gibt an, dass der strikte Modus aktiviert ist.	<code>false</code>

Einstellung	Beschreibung	Standard
<code>strict-mode-fail-query</code>	Weist Abfragen ab, wenn der strikte Modus Abfragen mit wahrscheinlich langer Laufzeit erkennt. Falls <code>false</code> , gibt Amazon EMR nur eine Warnung aus.	<code>false</code>
<code>strict-mode-restrictions</code>	Gibt die Einschränkungen an, die gelten sollen, wenn der strikte Modus aktiviert ist. Der strikte Modus unterstützt die folgenden Einschränkungen: <code>MANDATORY_PARTITION_PREDICATE</code> , <code>DISALLOW_CROSS_JOIN</code> und <code>LIMITED_SORT</code> .	<code>MANDATORY_PARTITION_PREDICATE</code> , <code>DISALLOW_CROSS_JOIN</code> , <code>LIMITED_SORT</code>

Um mit dem strikten Modus zu experimentieren, können Sie diese Konfigurationen überschreiben oder sie als Sitzungseigenschaften festlegen, wenn Sie den Presto-Client verwenden.

Um die Konfiguration bei der Clustererstellung festzulegen, verwenden Sie AWS Management Console

1. Wählen Sie Cluster erstellen und wählen Sie Amazon-EMR-Version 6.8.0 und Presto oder Trino aus. Weitere Informationen finden Sie unter [Installieren von PrestoDB und Trino](#).
2. Geben Sie die Konfigurationseigenschaften für den strikten Modus direkt an oder laden Sie eine JSON-Datei auf Amazon S3 hoch. Wählen Sie optional den für Ihren Metastore aus. Geben Sie Ihre VPC, Subnetze, Bootstrap-Aktionen, Ihr key pair und Ihre Sicherheitsgruppe an. Um Ihren Cluster jetzt zu erstellen, wählen Sie Cluster erstellen aus.
3. Melden Sie sich beim Primärknoten des Clusters an und führen Sie `presto-cli` oder `trino-cli` aus.
4. Reichen Sie Ihre Anfragen ein. Der strikte Modus validiert jede Abfrage und stellt fest, ob sie lange andauert. Abhängig von Ihrer `strict-mode-fail-query`-Einstellung lehnt Amazon EMR die Abfrage ab oder gibt eine Warnung aus.

5. Wenn Sie mit Ihren Abfragen fertig sind, beenden Sie den Cluster und löschen Sie Ihre Ressourcen.

Um die Konfiguration auf einem laufenden Cluster festzulegen mit dem AWS CLI

1. Melden Sie sich mit dem am primären Knoten Ihres Clusters an AWS CLI und führen Sie `presto-cli` oder `austrino-cli`.
2. Führen Sie die folgenden Befehle mit den gewünschten Werten aus.

```
set session strict_mode_enabled = true;  
set session strict_mode_fail_query = false;  
set session strict_mode_restrictions = 'DISALLOW_CROSS_JOIN,LIMITED_SORT';
```

## Überlegungen

Berücksichtigen Sie bei der Verwendung des strikten Modus die folgenden Punkte:

- In einigen Fällen kann der strikte Modus Abfragen mit kurzer Laufzeit zurückweisen, die nicht viele Ressourcen verbrauchen. Beispielsweise wenden Abfragen in kleinen Tabellen keine dynamische Filterung an und ersetzen auch keine inneren Verknüpfungen durch Kreuzverknüpfungen. Dies kann dazu führen, dass die Abfrage das obligatorische Partitionsprädikat verwendet oder Cross-Joins nicht zulässt. In diesem Fall lehnt der strikte Modus die Abfrage ab.
- Die Prüfung im strikten Modus wird nur auf die Abfragetypen SELECT, INSERT, CREATE TABLE AS SELECT und EXPLAIN ANALYZE angewendet.
- Sie können die Einschränkung für das obligatorische Partitionsprädikat nur mit dem Hive-Konnektor verwenden.

## Umgang mit dem Verlust von Spot Instances in Presto

Mit Spot Instances in Amazon EMR können Sie Big-Data-Workloads auf freien Amazon-EC2-Kapazitäten zu reduzierten Kosten ausführen. Als Gegenleistung für die geringeren Kosten kann Amazon-EC2-Spot Instances mit einer zweiminütigen Benachrichtigung unterbrechen. Wenn Sie einen Knoten beenden, kann es bis zu 10 Minuten dauern, bis Presto einen Fehler zurückgibt. Dies führt zu unnötigen Verzögerungen bei Fehlerberichten und möglichen Wiederholungsversuchen. Schnelle Beendigung ist ein Feature, mit der Sie kontrollieren können, wie Presto mit beendeten Knoten umgeht.

Die Aufgabe des Presto-Koordinators besteht darin, den Überblick über alle Worker-Knoten zu behalten und deren Status regelmäßig abzufragen. Ohne schnelle Terminierung konsultiert der Koordinator das YARN nicht, NodeManager um den Status der einzelnen Knoten zu erfahren. Dies kann zu einer langen Wiederholungsschleife führen, bevor die Abfrage fehlschlägt. Bei einer Schnellbeendigung fragt der Presto-Koordinator den Knotenstatus ab, NodeManager sobald die Abfrage den Host nicht erreicht. Wenn sich NodeManager herausstellt, dass der Knoten inaktiv ist, bricht Presto weitere Versuche ab, schlägt die Abfrage fehl und gibt einen Fehler zurück.

`NODE_DECOMMISSIONED`

Mit den folgenden Konfigurationsparametern können Sie das Verhalten von Presto im Falle einer Knotenbeendigung steuern und anpassen.

### Presto-Konfigurationen für die Behandlung von Knotenausfällen

Einstellung	Beschreibung	Standard
<code>query.remote-task.max-backoff-duration</code>	Die Dauer, während der der Koordinator weiterhin versucht, den Status der Remote-Aufgabe von den Worker-Knoten abzurufen.	10 Minuten
<code>query.remote-task.quick-terminate-on-node-failure</code>	Aktiviert den schnellen Knotenausfall, wenn der Koordinator den Knoten nicht erreichen oder keine Verbindung zu dem Worker herstellen kann, der auf diesem Knoten ausgeführt wird. Der Wert von <code>query.remote-task.terminate-on-connection-exception</code> bestimmt, ob der Koordinator den Knoten erreichen oder eine Verbindung zum Worker herstellen muss.	true

Einstellung	Beschreibung	Standard
	<p>Der Knoten schlägt bei der Abfrage fehl und Amazon EMR entfernt den Knoten aus der Liste der verfügbaren Worker. In diesem Fall können Sie den Knoten nicht verwenden, um neue Abfragen zu planen.</p> <p>Wenn Sie diesen Wert auf <code>false</code> setzen, kehrt Presto zu seinem vorherigen Verhalten zurück, bei dem der Presto-Koordinator erneut versucht, den Knoten (für <code>query.remote-task.max-backoff-duration</code>) zu erreichen, bevor er den Knoten als nicht verfügbar markiert und die laufende Abfrage auf dem Knoten fehlschlägt.</p>	
<code>query.remote-task.terminate-on-connection-exception</code>	<p>Gibt an, ob Amazon EMR ein Knoten sein soll, wenn der Host erreichbar ist, der Koordinator aber keine Verbindung zum Worker-Prozess des Hosts herstellen kann. Wenn Sie diesen Wert auf <code>true</code> setzen, aktivieren Sie „schneller Abfragefehler“, falls der Host nicht erreichbar ist.</p>	<code>false</code>

# Fehlertolerante Ausführung in Trino

Die fehlertolerante Ausführung ist ein Mechanismus in Trino, den ein Cluster verwenden kann, um Abfragefehler zu minimieren. Zu diesem Zweck werden Abfragen oder deren Komponentenaufgaben wiederholt, wenn sie fehlschlagen. Wenn die fehlertolerante Ausführung aktiviert ist, werden die zwischengeschalteten Austauschdaten gespoolt, sodass sie im Falle eines Worker-Ausfalls oder eines anderen Fehlers während der Abfrageausführung von einem anderen Worker wiederverwendet werden können.

Weitere Informationen zur fehlertoleranten Ausführung in Trino finden Sie im Trino-Blog unter [Project Tardigrade bietet Anfängern ETL mit Trino-Geschwindigkeiten](#).

## Konfiguration

Die fehlertolerante Ausführung ist standardmäßig deaktiviert. Um das Feature zu aktivieren, legen Sie die `retry-policy` Konfigurationseigenschaft in der `trino-config`-Klassifizierung wie folgt auf `QUERY` oder `TASK` basierend auf der gewünschten Wiederholungsrichtlinie fest.

```
{
  "classification":
    "trino-config",
  "properties":
    {
      "retry-policy":
        "QUERY"
    }
}
```

Eine **QUERY**Wiederholungsrichtlinie weist Trino an, eine Abfrage automatisch zu wiederholen, wenn auf einem Worker-Knoten ein Fehler auftritt. Wir empfehlen, eine `QUERY` Wiederholungsrichtlinie zu verwenden, wenn der Großteil der Workload für den Trino-Cluster aus vielen kleinen Abfragen besteht.

Eine **TASK**Wiederholungsrichtlinie weist Trino an, einzelne Abfrageaufgaben im Falle eines Fehlers erneut zu versuchen. Wir empfehlen diese Richtlinie, wenn Trino umfangreiche Batch-Abfragen ausführt. Der Cluster kann kleinere Aufgaben innerhalb der Abfrage effizienter wiederholen, als die gesamte Abfrage erneut zu versuchen.

## Austausch-Manager

Ein Exchange-Manager speichert und verwaltet Spooldaten für eine fehlertolerante Ausführung. Er verwendet externen Speicher, um verschüttete Daten zu speichern, die über die Größe des In-Memory-Puffers hinausgehen. Sie können einen dateisystembasierten Exchange-Manager konfigurieren, der Spooldaten an einem bestimmten Ort speichert, z. B. Amazon S3, Amazon-S3-kompatible Systeme oder HDFS.

Die Amazon-EMR-Versionen 6.9.0 und höher enthalten die `trino-exchange-manager`-Klassifizierung zur Konfiguration des Exchange Managers. Diese Versionen unterstützen auch HDFS für das Spooling.

### Exchange Manager einrichten

Verwenden Sie die `trino-exchange-manager`-Konfigurationsklassifizierung, um einen Exchange Manager zu konfigurieren. Diese Klassifizierung erstellt intern eine `etc/exchange-manager.properties`-Konfigurationsdatei für den Koordinator und alle Worker-Knoten. Verwenden Sie die Konfigurationsklassifizierung , um die `exchange-manager.name`-Eigenschaften auf `filesystem` festzulegen.

Standardmäßig verwenden Amazon-EMR-Versionen 6.9.0 und höher HDFS als Exchange-Manager. HDFS ist in den Amazon-EMR-EC2-Clustern verfügbar, und das Spooling erfolgt standardmäßig im `trino-exchange/-`Verzeichnis. Um die Standardeinstellungen zu verwenden, legen Sie die folgende Konfiguration fest:

```
{"Classification":  
  "trino-exchange-manager"  
}
```

Wenn Sie einen benutzerdefinierten Speicherort angeben möchten, legen Sie die folgenden Eigenschaften in der `trino-exchange-manager`-Klassifizierung fest:

- Setzen Sie `exchange.use-local-hdfs` auf `true`.
- Legen Sie `exchange.base-directories` fest, um den Speicherort des benutzerdefinierten Verzeichnisses in HDFS darzustellen, z. B. `exchange.base-directories=/exchange`. Wenn sich das benutzerdefinierte Verzeichnis noch nicht in HDFS befindet, wird es von Amazon EMR erstellt.

### HDFS-Exchange-Manager-Konfigurationen



Basierend auf internen Testergebnissen empfehlen wir, auf lokales HDFS zu spoolen, um im Vergleich zu anderen cloudbasierten Dateisystemen eine bessere Abfrageleistung zu erzielen. Sie können die folgenden Konfigurationen für den Exchange-Manager mit HDFS festlegen.

Konfiguration	Beschreibung	Standardeinstellung
<code>exchange.hdfs.block-size</code>	Blockgröße für HDFS-Speicher	4 MB
<code>hdfs.config.resources</code>	Liste der Dateipfade zur Konfiguration von HDFS	Falls <code>exchange.use-local-hdfs true</code> ist, werden die Pfade zu <code>core-site.xml</code> , <code>hdfs-site.xml</code> Dateien verwendet, andernfalls <code>null</code>

Zusätzliche Konfigurationseigenschaften für die fehlertolerante Ausführung und Informationen zur Einrichtung von Amazon S3 oder anderen Amazon-S3-kompatiblen Systemen für das Spooling finden Sie auf der Seite [Fehlertolerante Ausführung](#) in der Trino-Dokumentation.

## Überlegungen und Einschränkungen

- Wenn Sie die fehlertolerante Ausführung aktivieren, werden `write`-Operationen für Konnektoren deaktiviert, die `write` nicht unterstützen, wenn `retry-policy` festgelegt ist. Ab der Amazon-EMR-Version 6.9.0 unterstützen Delta Lake-, Hive- und Iceberg-Konnektoren `write`-Vorgänge mit `retry-policy`.
- Wenn Sie Exchange Manager verwenden und teure E/A-Vorgänge durchführen, kann es bei Ihren Abfragen zu Leistungseinbußen kommen, während Exchange Manager die Zwischendaten auf externen Speicher überträgt.

## Auto Scaling von Presto mit ordnungsgemäßer Stilllegung verwenden

Die Amazon-EMR-Release-Versionen 5.30.0 und höher enthalten ein Feature, mit dem Sie einen Kulanzzzeitraum für bestimmte Skalierungsaktionen festlegen können. Die Übergangszeit ermöglicht es, dass Presto-Aufgaben weiterlaufen, bevor der Knoten aufgrund einer Skalierungsaktion zur

Größenänderung oder einer automatischen Skalierungsrichtlinienanforderung beendet wird. Weitere Informationen zu Skalierungsregeln finden Sie unter [Grundlegendes zu automatischen Skalierungsregeln](#) im Verwaltungshandbuch für Amazon EMR. Auto Scaling in Presto mit ordnungsgemäßer Stilllegung verhindert, dass neue Aufgaben auf einem Knoten geplant werden, der außer Betrieb genommen wird, und ermöglicht gleichzeitig, dass bereits laufende Aufgaben abgeschlossen werden, bevor die Zeitüberschreitung für das Herunterfahren erreicht ist. Derzeit laufende Abfragen schließen die Ausführung ab, bevor der Knoten stillgelegt wird. Auto Scaling wird für Instance-Flotten nicht unterstützt.

Sie können steuern, wie viel Zeit für die Ausführung von Presto-Aufgaben eingeräumt werden soll, nachdem eine Anfrage zum Herunterfahren der automatischen Skalierung eingegangen ist. Standardmäßig ist das Timeout für das Herunterfahren für Amazon EMR 0 Minuten, wodurch der Knoten und alle darauf ausgeführten Presto-Aufgaben sofort beendet werden, falls dies gemäß einer Abskalierungs-Anforderung erforderlich ist. Um ein längeres Timeout für Presto-Aufgaben auf Amazon EMR festzulegen, damit laufende Abfragen vor der Herunterskalierung eines Clusters abgeschlossen werden können, legen Sie über die `presto-config`-Konfigurationsklassifizierung den Parameter `graceful-shutdown-timeout` auf einen Wert in Sekunden oder Minuten fest, der größer als Null ist. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Wenn Sie beispielsweise den Wert `graceful-shutdown-timeout` auf "30m" erhöhen, wird ein Zeitraum von 30 Minuten angegeben. Nach Ablauf der Zeitüberschreitung für das Herunterfahren wird der Knoten, der für die Stilllegung markiert ist, zwangsweise beendet, wenn er auf den Abschluss von Abfrageaufgaben wartet und die Abfrage fehlschlägt. Wenn die Abfrageaufgaben in 5 Minuten abgeschlossen sind, wird der für die Stilllegung markierte Knoten nach fünf Minuten beendet, sofern andere YARN Anwendungen die Ausführung abgeschlossen haben.

Example Beispiel einer automatischen Skalierung in Presto mit ordnungsgemäßer Stilllegung

Ersetzen Sie den `graceful-shutdown-timeout`-Wert durch die Anzahl der Minuten, die für Ihre Einrichtung geeignet ist. Es gibt keinen Höchstwert. Im folgenden Beispiel wird ein Timeout-Wert von 1800 Sekunden (30 Minuten) festgelegt.

```
[
  {
    "classification": "presto-config",
    "properties": {
      "graceful-shutdown-timeout": "1800s"
    }
  }
]
```

## Einschränkungen

Ordnungsgemäße Stilllegung in PrestoDB funktioniert nicht auf EMR-Clustern, auf denen die HTTP-Konnektivität deaktiviert ist, z. B. wenn `http-server.http.enabled` auf `false` eingestellt ist. Trino unterstützt Ordnungsgemäße Stilllegung, unabhängig von der Einstellung `http-server.http.enabled`, überhaupt nicht.

## Überlegungen zu Presto in Amazon EMR

Berücksichtigen Sie die folgenden Unterschiede und Einschränkungen, wenn Sie [Presto](#) in Amazon EMR ausführen.

### In der Befehlszeile von Presto ausführbare Datei

In Amazon EMR verwenden PrestoDB und Trino dieselbe ausführbare Befehlszeilendatei `presto-cli`, wie im folgenden Beispiel.

```
presto-cli --catalog hive
```

### Nicht konfigurierbare Presto-Bereitstellungseigenschaften

Die Version von Amazon EMR, die Sie verwenden, bestimmt die verfügbaren Presto-Bereitstellungskonfigurationen. Weitere Informationen zu diesen Konfigurationseigenschaften finden Sie unter [Bereitstellen von Presto](#) in der Presto-Dokumentation. Die folgende Tabelle zeigt die Konfigurationsoptionen der Presto-properties-Dateien.

Datei	Konfigurierbar
<code>log.properties</code>	<p>PrestoDB: Konfigurierbar in Amazon-EMR-Versionen 4.0.0 und höher. Verwenden Sie die <code>presto-log</code> -Konfigurationsklassifizierung.</p> <p>Trino (PrestoSQL): Konfigurierbar in Amazon-EMR-Versionen 6.1.0 und höher. Verwenden Sie die <code>prestoql-log</code> - oder <code>trino-log</code> -Konfigurationsklassifizierung.</p>

Datei	Konfigurierbar
config.properties	<p>PrestoDB: Konfigurierbar in Amazon-EMR-Versionen 4.0.0 und höher. Verwenden Sie die <code>presto-config</code> -Konfigurationsklassifizierung.</p> <p>Trino (PrestoSQL): Konfigurierbar in Amazon-EMR-Versionen 6.1.0 und höher. Verwenden Sie die <code>prestoql-config</code> - oder <code>trino-config</code> -Konfigurationsklassifizierung.</p>
hive.properties	<p>PrestoDB: Konfigurierbar in der Amazon-EMR-Version 4.1.0 und höher. Verwenden Sie die <code>presto-connector-hive</code> -Konfigurationsklassifizierung.</p> <p>Trino (PrestoSQL): Konfigurierbar in Amazon-EMR-Versionen 6.1.0 und höher. Verwenden Sie die <code>prestoql-connector-hive</code> - oder <code>trino-connector-hive</code> -Konfigurationsklassifizierung.</p>
node.properties	<p>PrestoDB: Konfigurierbar in der Amazon-EMR-Version 5.6.0 und höher. Verwenden Sie die <code>presto-node</code> -Konfigurationsklassifizierung.</p> <p>Trino (PrestoSQL): Konfigurierbar in Amazon-EMR-Versionen 6.1.0 und höher. Verwenden Sie die <code>prestoql-node</code> - oder <code>trino-node</code> -Konfigurationsklassifizierung.</p>
jvm.config	Nicht konfigurierbar.

## Installation von PrestoDB und Trino

Der Anwendungsname Presto wird weiterhin verwendet, um PrestoDB auf Clustern zu installieren. Um Trino auf Clustern zu installieren, verwenden Sie den Anwendungsnamen Trino oder PrestoSQL in älteren Versionen von Amazon EMR).

Sie können entweder PrestoDB oder Trino installieren, aber Sie können nicht beide auf einem einzigen Cluster installieren. Wenn Sie beim Versuch, einen Cluster zu erstellen, sowohl PrestoDB als auch Trino angeben, tritt ein Validierungsfehler auf und die Anfrage zur Clustererstellung schlägt fehl.

## EMRFS- und PrestoS3-Konfiguration FileSystem

Mit Amazon-EMR-Versionen 5.12.0 und höher kann PrestoDB EMRFS verwenden. Dies ist die Standardkonfiguration. EMRFS ist auch das Standarddateisystem für Trino (PrestoSQL) in Amazon-EMR-Versionen 6.1.0 und höher. Weitere Informationen finden Sie unter [EMR File System \(EMRFS\)](#) im Verwaltungshandbuch für Amazon EMR. In früheren Versionen von Amazon EMR FileSystem ist PrestoS3 die einzige Konfigurationsoption.

Sie können eine Sicherheitskonfiguration verwenden, um die Verschlüsselung für EMRFS-Daten in Amazon S3 einzurichten. Sie können auch IAM-Rollen für EMRFS-Anforderungen an Amazon S3 verwenden. Weitere Informationen finden unter [Verschlüsselungsoptionen verstehen](#) und [IAM-Rollen für EMRFS in Amazon S3 konfigurieren](#) im Verwaltungshandbuch für Amazon EMR.

### Note

Wenn Sie die zugrunde liegenden Daten in Amazon S3 mit Amazon-EMR-Version 5.12.0 abfragen, können Presto-Fehler auftreten. Dies liegt daran, dass Presto die Konfigurationsklassifizierungswerte von `emrfs-site.xml` nicht abrufen. Um dieses Problem zu umgehen, erstellen Sie ein `emrfs`-Unterverzeichnis unter `usr/lib/presto/plugin/hive-hadoop2/` und erstellen einen Symlink `usr/lib/presto/plugin/hive-hadoop2/emrfs` zu der vorhandenen Datei `/usr/share/aws/emr/emrfs/conf/emrfs-site.xml`. Starten Sie dann den Presto-Server-Prozess neu (`sudo presto-server stop` gefolgt von `sudo presto-server start`).

Sie können den EMRFS-Standard überschreiben und stattdessen PrestoS3 verwenden. FileSystem Verwenden Sie dazu die `presto-connector-hive`-Konfigurationsklassifizierung, um `hive.s3-`

`file-system-type` auf `PRESTO` zu setzen, wie im folgenden Beispiel veranschaulicht. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

```
[
  {
    "Classification": "presto-connector-hive",
    "Properties": {
      "hive.s3-file-system-type": "PRESTO"
    }
  }
]
```

Wenn Sie PrestoS3 verwenden, verwenden Sie die `presto-connector-hive` Konfigurationsklassifizierung oder `trino-connector-hive` für TrinoFileSystem, um PrestoS3-Eigenschaften zu konfigurieren. Weitere Informationen zu den verfügbaren Eigenschaften finden Sie unter [Amazon-S3-Konfiguration](#) im Abschnitt zum Hive-Konnektor in der Presto-Dokumentation. Diese Einstellungen gelten nicht für EMRFS.

## Standardeinstellung für Endbenutzer-Identitätswechsel

Standardmäßig ist der Endbenutzer-Identitätswechsel ab der Amazon-EMR-Version 5.12.0 und höher für den Zugriff auf HDFS aktiviert. Weitere Informationen finden Sie im Abschnitt zum [Endbenutzer-Identitätswechsel](#) in der Presto-Dokumentation. Um diese Einstellung zu ändern, setzen Sie die `presto-config`-Konfigurationsklassifizierung der `hive.hdfs.impersonation.enabled`-Eigenschaft auf `false`.

## Standardport für die Presto-Webschnittstelle

Standardmäßig konfiguriert Amazon EMR die Presto-Webschnittstelle auf dem Presto-Koordinator dazu, den Port 8889 (für PrestoDB und Trino) zu verwenden. Um den Port zu ändern, verwenden Sie die `presto-config`-Konfigurationsklassifizierung, um die `http-server.http.port`-Eigenschaft festzulegen. Weitere Informationen finden Sie unter [Konfigurationseigenschaften](#) im Abschnitt zur Bereitstellung von Presto in der Presto-Dokumentation.

## Problem bei der Hive-Bucket-Ausführung in einigen Versionen

Presto Version 152.3 hat ein Problem mit der Hive-Bucket-Ausführung, das in Presto in einigen Fällen zu einer erheblich langsameren Abfrageleistung führt. Die Amazon-EMR-Versionen 5.0.3, 5.1.0 und 5.2.0 enthalten diese Version von Presto. Um dieses Problem abzumildern, verwenden Sie

die `presto-connector-hive`-Konfigurationsklassifizierung, um die Eigenschaft `hive.bucket-execution` auf `false` zu setzen, wie im folgenden Beispiel veranschaulicht.

```
[
  {
    "Classification": "presto-connector-hive",
    "Properties": {
      "hive.bucket-execution": "false"
    }
  }
]
```

## Presto-Versionsverlauf

In der folgenden Tabelle sind die Version von Presto aufgeführt, die in jeder Release-Version von Amazon EMR enthalten ist, zusammen mit den Komponenten, die mit der Anwendung installiert wurden. Informationen zu den Komponentenversionen in den einzelnen Versionen finden Sie im Abschnitt [Komponentenversion für Ihre Version in Amazon-EMR-7.x-Versionen](#), [Amazon-EMR-6.x-Versionen](#) oder [Amazon-EMR-5.x-Versionen](#).

Die Presto-Versionsinformationen

Amazon-EMR-Version sbezeichnung	Presto-Version	Mit Presto installierte Komponenten
emr-5.36.2	0,267	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Amazon-EMR-Version sbezeichnung	Presto-Version	Mit Presto installierte Komponenten
emr-7.1.0	0,284	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-7.0.0	0,283	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker



Amazon-EMR-Version sbezeichnung	Presto-Version	Mit Presto installierte Komponenten
emr-6.15.0	0,283	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-6.14.0	0,281	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Amazon-EMR-Version sbezeichnung	Presto-Version	Mit Presto installierte Komponenten
emr-6.13.0	0,281	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-6.12.0	0,281	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Amazon-EMR-Version sbezeichnung	Presto-Version	Mit Presto installierte Komponenten
emr-6.11.1	0,279	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-6.11.0	0,279	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Amazon-EMR-Version sbezeichnung	Presto-Version	Mit Presto installierte Komponenten
emr-6.10.1	0,278	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-6.10.0	0,278	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Amazon-EMR-Version sbezeichnung	Presto-Version	Mit Presto installierte Komponenten
emr-6.9.1	0,276	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-6.9.0	0,276	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Amazon-EMR-Version sbezeichnung	Presto-Version	Mit Presto installierte Komponenten
emr-6.8.1	0,273	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-6.8.0	0,273	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Amazon-EMR-Version sbezeichnung	Presto-Version	Mit Presto installierte Komponenten
emr-6.7.0	0,272	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-5.36.1	0,267	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Amazon-EMR-Version sbezeichnung	Presto-Version	Mit Presto installierte Komponenten
emr-5.36.0	0,267	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-6.6.0	0,267	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker



Amazon-EMR-Version sbezeichnung	Presto-Version	Mit Presto installierte Komponenten
emr-5.35.0	0,266	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-6.5.0	0,261	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Amazon-EMR-Version sbezeichnung	Presto-Version	Mit Presto installierte Komponenten
emr-6.4.0	0,254,1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-6.3.1	0,245,1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Amazon-EMR-Version sbezeichnung	Presto-Version	Mit Presto installierte Komponenten
emr-6.3.0	0,245,1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-6.2.1	0,238,3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Amazon-EMR-Version sbezeichnung	Presto-Version	Mit Presto installierte Komponenten
emr-6.2.0	0,238,3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-6.1.1	0.232	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Amazon-EMR-Version sbezeichnung	Presto-Version	Mit Presto installierte Komponenten
emr-6.1.0	0.232	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-6.0.1	0.230	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Amazon-EMR-Version sbezeichnung	Presto-Version	Mit Presto installierte Komponenten
emr-6.0.0	0.230	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-5.34.0	0,261	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Amazon-EMR-Version sbezeichnung	Presto-Version	Mit Presto installierte Komponenten
emr-5.33.1	0,245,1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-5.33.0	0,245,1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Amazon-EMR-Version sbezeichnung	Presto-Version	Mit Presto installierte Komponenten
emr-5.32.1	0,240,1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-5.32.0	0,240,1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker



Amazon-EMR-Version sbezeichnung	Presto-Version	Mit Presto installierte Komponenten
emr-5.31.1	0,238,3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-5.31.0	0,238,3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Amazon-EMR-Version sbezeichnung	Presto-Version	Mit Presto installierte Komponenten
emr-5.30.2	0.232	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-5.30.1	0.232	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Amazon-EMR-Version sbezeichnung	Presto-Version	Mit Presto installierte Komponenten
emr-5.30.0	0.232	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-5.29.0	0.227	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon-EMR-Version sbezeichnung	Presto-Version	Mit Presto installierte Komponenten
emr-5.28.1	0.227	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.28.0	0.227	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon-EMR-Version sbezeichnung	Presto-Version	Mit Presto installierte Komponenten
emr-5.27.1	0.224	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.27.0	0.224	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon-EMR-Version sbezeichnung	Presto-Version	Mit Presto installierte Komponenten
emr-5.26.0	0.220	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.25.0	0.220	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon-EMR-Version sbezeichnung	Presto-Version	Mit Presto installierte Komponenten
emr-5.24.1	0.219	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.24.0	0.219	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon-EMR-Version sbezeichnung	Presto-Version	Mit Presto installierte Komponenten
emr-5.23.1	0.215	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.23.0	0.215	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker



Amazon-EMR-Version sbezeichnung	Presto-Version	Mit Presto installierte Komponenten
emr-5.22.0	0.215	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.21.2	0.215	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon-EMR-Version sbezeichnung	Presto-Version	Mit Presto installierte Komponenten
emr-5.21.1	0.215	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.21.0	0.215	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon-EMR-Version sbezeichnung	Presto-Version	Mit Presto installierte Komponenten
emr-5.20.1	0.214	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.20.0	0.214	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon-EMR-Version sbezeichnung	Presto-Version	Mit Presto installierte Komponenten
emr-5.19.1	0.212	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.19.0	0.212	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon-EMR-Version sbezeichnung	Presto-Version	Mit Presto installierte Komponenten
emr-5.18.1	0.210	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.18.0	0.210	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon-EMR-Version sbezeichnung	Presto-Version	Mit Presto installierte Komponenten
emr-5.17.2	0.206	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.17.1	0.206	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon-EMR-Version sbezeichnung	Presto-Version	Mit Presto installierte Komponenten
emr-5.17.0	0.206	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.16.1	0.203	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon-EMR-Version sbezeichnung	Presto-Version	Mit Presto installierte Komponenten
emr-5.16.0	0.203	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.15.1	0.194	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker



Amazon-EMR-Version sbezeichnung	Presto-Version	Mit Presto installierte Komponenten
emr-5.15.0	0.194	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.14.2	0.194	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon-EMR-Version sbezeichnung	Presto-Version	Mit Presto installierte Komponenten
emr-5.14.1	0.194	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.14.0	0.194	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon-EMR-Version sbezeichnung	Presto-Version	Mit Presto installierte Komponenten
emr-5.13.1	0.194	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.13.0	0.194	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon-EMR-Version sbezeichnung	Presto-Version	Mit Presto installierte Komponenten
emr-5.12.3	0.188	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.12.2	0.188	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon-EMR-Version sbezeichnung	Presto-Version	Mit Presto installierte Komponenten
emr-5.12.1	0.188	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.12.0	0.188	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon-EMR-Version sbezeichnung	Presto-Version	Mit Presto installierte Komponenten
emr-5.11.4	0.187	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.11.3	0.187	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon-EMR-Version sbezeichnung	Presto-Version	Mit Presto installierte Komponenten
emr-5.11.2	0.187	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.11.1	0.187	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon-EMR-Version sbezeichnung	Presto-Version	Mit Presto installierte Komponenten
emr-5.11.0	0.187	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.10.1	0.187	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker



Amazon-EMR-Version sbezeichnung	Presto-Version	Mit Presto installierte Komponenten
emr-5.10.0	0.187	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.9.1	0.184	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon-EMR-Version sbezeichnung	Presto-Version	Mit Presto installierte Komponenten
emr-5.9.0	0.184	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.8.3	0.170	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon-EMR-Version sbezeichnung	Presto-Version	Mit Presto installierte Komponenten
emr-5.8.2	0.170	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.8.1	0.170	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon-EMR-Version sbezeichnung	Presto-Version	Mit Presto installierte Komponenten
emr-5.8.0	0.170	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.7.1	0.170	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon-EMR-Version sbezeichnung	Presto-Version	Mit Presto installierte Komponenten
emr-5.7.0	0.170	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.6.1	0.170	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon-EMR-Version sbezeichnung	Presto-Version	Mit Presto installierte Komponenten
emr-5.6.0	0.170	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.5.4	0.170	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.5.3	0.170	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon-EMR-Version sbezeichnung	Presto-Version	Mit Presto installierte Komponenten
emr-5.5.2	0.170	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.5.1	0.170	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.5.0	0.170	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon-EMR-Version sbezeichnung	Presto-Version	Mit Presto installierte Komponenten
emr-5.4.1	0.166	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.4.0	0.166	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.3.2	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker



Amazon-EMR-Version sbezeichnung	Presto-Version	Mit Presto installierte Komponenten
emr-5.3.1	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.3.0	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.2.3	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon-EMR-Version sbezeichnung	Presto-Version	Mit Presto installierte Komponenten
emr-5.2.2	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.2.1	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.2.0	0.152.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon-EMR-Version sbezeichnung	Presto-Version	Mit Presto installierte Komponenten
emr-5.1.1	0.152.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.1.0	0.152.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.0.3	0.152.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon-EMR-Version sbezeichnung	Presto-Version	Mit Presto installierte Komponenten
emr-5.0.2	0.150	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.0.1	0.150	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.0.0	0.150	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

In der folgenden Tabelle sind die Versionen von Trino (Presto SQL) aufgeführt, die in jeder Release-Version von Amazon EMR enthalten ist, zusammen mit den Komponenten, die mit der Anwendung installiert wurden. PrestoSQL änderte seinen Namen ab Version 351 in Trino.

### Trino (PrestoSQL)-Versionsinformationen

Amazon-EMR-Version sbezeichnung	Trino (PrestoSQL)-Version	Mit Trino (PrestoSQL) installierte Komponenten
emr-7.1.0	435	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker
emr-7.0.0	426	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker
emr-6.15.0	426	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodema

Amazon-EMR-Version sbezeichnung	Trino (PrestoSQL)-Version	Mit Trino (PrestoSQL) installierte Komponenten
		nager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker
emr-6.14.0	422	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker
emr-6.13.0	414	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker

Amazon-EMR-Version sbezeichnung	Trino (PrestoSQL)-Version	Mit Trino (PrestoSQL) installierte Komponenten
emr-6.12.0	414	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker
emr-6.11.1	410	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker

Amazon-EMR-Version sbezeichnung	Trino (PrestoSQL)-Version	Mit Trino (PrestoSQL) installierte Komponenten
emr-6.11.0	410	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker
emr-6.10.1	403	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker



Amazon-EMR-Version sbezeichnung	Trino (PrestoSQL)-Version	Mit Trino (PrestoSQL) installierte Komponenten
emr-6.10.0	403	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker
emr-6.9.1	398	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker

Amazon-EMR-Version sbezeichnung	Trino (PrestoSQL)-Version	Mit Trino (PrestoSQL) installierte Komponenten
emr-6.9.0	398	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker
emr-6.8.1	388	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker

Amazon-EMR-Version sbezeichnung	Trino (PrestoSQL)-Version	Mit Trino (PrestoSQL) installierte Komponenten
emr-6.8.0	388	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker
emr-6.7.0	378	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker

Amazon-EMR-Version sbezeichnung	Trino (PrestoSQL)-Version	Mit Trino (PrestoSQL) installierte Komponenten
emr-6.6.0	367	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker
emr-6.5.0	360	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker

Amazon-EMR-Version sbezeichnung	Trino (PrestoSQL)-Version	Mit Trino (PrestoSQL) installierte Komponenten
emr-6.4.0	359	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker
emr-6.3.1	350	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-prestosql, hcatalog-server, mariadb-server, prestosql-coordinator, prestosql-worker

Amazon-EMR-Version sbezeichnung	Trino (PrestoSQL)-Version	Mit Trino (PrestoSQL) installierte Komponenten
emr-6.3.0	350	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-prestosql, hcatalog-server, mariadb-server, prestosql-coordinator, prestosql-worker
emr-6.2.1	343	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-prestosql, hcatalog-server, mariadb-server, prestosql-coordinator, prestosql-worker

Amazon-EMR-Version sbezeichnung	Trino (PrestoSQL)-Version	Mit Trino (PrestoSQL) installierte Komponenten
emr-6.2.0	343	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-prestosql, hcatalog-server, mariadb-server, prestosql-coordinator, prestosql-worker
emr-6.1.1	338	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-prestosql, hcatalog-server, mariadb-server, prestosql-coordinator, prestosql-worker

Amazon-EMR-Version sbezeichnung	Trino (PrestoSQL)-Version	Mit Trino (PrestoSQL) installierte Komponenten
emr-6.1.0	338	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-prestosql, hcatalog-server, mariadb-server, prestosql-coordinator, prestosql-worker

## Trino (PrestoSQL) Versionshinweise nach Version

- [Amazon EMR 6.9.0 – Versionshinweise zu Trino \(PrestoSQL\)](#)

### Amazon EMR 6.9.0 – Versionshinweise zu Trino (PrestoSQL)

#### Amazon EMR 6.9.0 – Neue Feature von Trino (PrestoSQL)

- Um Abfragen mit langer Laufzeit zu unterstützen, verfügt Trino jetzt über einen fehlertoleranten Ausführungsmechanismus. Die fehlertolerante Ausführung minimiert Abfragefehler, indem fehlgeschlagene Abfragen oder deren Komponentenaufgaben wiederholt werden. Weitere Informationen finden Sie unter [Fehlertolerante Ausführung in Trino](#).

#### Amazon EMR 6.9.0 – Änderungen an Trino (PrestoSQL)

#### Amazon EMR 6.9.0 – PrestoDB-Änderungen

Typ	Beschreibung
Upgrade	Upgrade auf PrestoDB 0.276



Typ	Beschreibung
Upgrade	Unterstützung von Hadoop 3.3.3
Upgrade	Hudi-Upgrade auf 0.12.1
Funktion	Integration von Amazon EMR und Presto AWS Lake Formation für interaktive Workloads mithilfe der GCSC-API.
Funktion	Kerberos-bezogene Konfigurationen wurden in Sicherheitskonfigurationen für PrestoDB hinzugefügt, um Kerberos zu aktivieren.
Fehlerbehebung	Die OSS-Pull-Anfrage #18115, die hinzugefügt wurde, um die Anzahl der <code>hdfsConfiguration</code> -Kopien zu reduzieren, wurde rückgängig gemacht. Dies führte dazu, dass die HDFS-Konfiguration falsch kopiert wurde, wenn EMRFS- oder Hudi-Tabellen verwendet wurden.

## Amazon EMR 6.9.0 – Trino-Änderungen

Typ	Beschreibung
Upgrade	Trino-Upgrade auf 398
Upgrade	Unterstützung von Hadoop 3.3.3
Funktion	Tardigrade-Unterstützung: Unterstützung für Exchange-Spooling auf HDFS und Amazon S3 hinzugefügt. Weitere Informationen finden Sie unter <a href="#">Fehlertolerante Ausführung in Trino</a> .
Fehlerbehebung	Wenn Trino Iceberg verwendet wird und der Glue-Katalog aktiviert ist, vermeiden Sie das

Typ	Beschreibung
	Hinzufügen von Metastore-URI in <code>iceberg.properties</code> .

## Amazon EMR 6.9.0 – Bekannte Probleme mit Trino (PrestoSQL)

- Für Amazon-EMR-Version 6.9.0 funktioniert Trino nicht auf Clustern, die für Apache Ranger aktiviert sind. Wenn Sie Trino mit Ranger verwenden müssen, wenden Sie sich an [AWS Support](#).

# Apache Spark

[Apache Spark](#) ist ein verteiltes Verarbeitungs-Framework und Programmiermodell, mit dem Sie Machine Learning, Stream-Verarbeitung oder Graph-Analysen mit Amazon-EMR-Clustern durchführen können. Ähnlich wie Apache Hadoop ist Spark ein verteiltes Open-Source-Verarbeitungssystem, das häufig für Big-Data-Workloads verwendet wird. Spark weist jedoch einige bemerkenswerte Unterschiede zu Hadoop MapReduce auf. Spark hat eine optimierte DAG-Ausführungs-Engine (Directed Acyclic Graph, gerichteter azyklischer Graph) und betreibt aktives In-Memory-Caching für Daten. Dies kann die Leistung insbesondere für bestimmte Algorithmen und interaktive Abfragen steigern.

Spark unterstützt standardmäßig Anwendungen, die in Scala, Java und Python geschrieben sind. Außerdem enthält es mehrere hochgradig ineinander integrierte Bibliotheken für SQL ([Spark SQL](#)), Machine Learning ([MLlib](#)), Stream-Verarbeitung ([Spark Streaming](#)) und Graph-Verarbeitung ([GraphX](#)). Diese Tools vereinfachen die Nutzung des Spark-Frameworks für eine Vielzahl von Anwendungsfällen.

Sie können Spark zusammen mit anderen Hadoop-Anwendungen auf einem Amazon-EMR-Cluster installieren. Es kann außerdem das Amazon-EMR-Dateisystem (EMRFS) nutzen, um direkt auf Daten in Amazon S3 zuzugreifen. Hive ist auch in Spark integriert, sodass Sie ein HiveContext Objekt verwenden können, um Hive-Skripte mit Spark auszuführen. Ein Hive-Kontext ist als `sqlContext` Bestandteil der Spark-Shell.

Ein Beispiel-Tutorial zur Einrichtung eines EMR-Clusters mit Spark und zur Analyse eines Beispieldatensatzes finden Sie unter [Tutorial: Erste Schritte mit Amazon EMR](#) im AWS News-Blog.

## Important

Apache-Spark-Version 2.3.1, verfügbar ab Amazon-EMR-Version 5.16.0, adressiert [CVE-2018-8024](#) und [CVE-2018-1334](#). Wir empfehlen, dass Sie frühere Versionen von Spark zu Spark-Version 2.3.1 oder höher migrieren.

Die folgende Tabelle listet die Version von Spark auf, die in der neuesten Version der Amazon-EMR-7.x-Serie enthalten ist, zusammen mit den Komponenten, die Amazon EMR mit Spark installiert.

Informationen zur Version der Komponenten, die in dieser Version mit Spark installiert wurden, finden Sie unter [Komponentenversionen von Version 7.1.0](#).

#### Spark-Versionsinformationen für emr-7.1.0

Amazon-EMR-Version sbezeichnung	Spark-Version	Mit Spark installierte Komponenten
emr-7.1.0	Spark 3.5.0	delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave


Die folgende Tabelle listet die Version von Spark auf, die in der neuesten Version der Amazon-EMR-6.x-Serie enthalten ist, zusammen mit den Komponenten, die Amazon EMR mit Spark installiert.

Die Version der Komponenten, die mit Spark in dieser Version installiert wurden, finden Sie unter [Komponentenversionen der Version 6.15.0](#).

#### Spark-Versionsinformationen für emr-6.15.0

Amazon-EMR-Version sbezeichnung	Spark-Version	Mit Spark installierte Komponenten
emr-6.15.0	Spark 3.4.1	aws-sagemaker-spark-sdk, delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-

Amazon-EMR-Version sbezeichnung	Spark-Version	Mit Spark installierte Komponenten
		datanode, hadoop-hdfs-librar y, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, hudi, hudi- spark, iceberg, livy-server, nginx, r, spark-client, spark- history-server, spark-on-yarn, spark-yarn-slave

 Note

Amazon-EMR-Version 6.8.0 wird mit Apache Spark 3.3.0 geliefert. Diese Spark-Version verwendet Apache Log4j 2 und die `log4j2.properties`-Datei zur Konfiguration von Log4j in Spark-Prozessen. Wenn Sie Spark im Cluster verwenden oder EMR-Cluster mit benutzerdefinierten Konfigurationsparametern erstellen und ein Upgrade auf Amazon-EMR-Version 6.8.0 durchführen möchten, müssen Sie auf die neue `spark-log4j2`-Konfigurationsklassifizierung und das neue Schlüsselformat für Apache Log4j 2 migrieren. Weitere Informationen finden Sie unter [Migration von Apache Log4j 1.x zu Log4j 2.x](#).

Die folgende Tabelle listet die Version von Spark auf, die in der neuesten Version der Amazon-EMR-5.x-Serie enthalten ist, zusammen mit den Komponenten, die Amazon EMR mit Spark installiert.

Informationen zur Version der Komponenten, die in dieser Version mit Spark installiert wurden, finden Sie unter [Komponentenversionen von Version 5.36.2](#).

## Spark-Versionsinformationen für emr-5.36.2

Amazon-EMR-Version sbezeichnung	Spark-Version	Mit Spark installierte Komponenten
emr-5.36.2	Spark 2.4.8	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr- ddb, emr-s3-select, hadoop-cl ient, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server, hudi, hudi- spark, livy-server, nginx, r, spark-client, spark-history-serv er, spark-on-yarn, spark-yarn- slave

## Themen

- [Einen Cluster mit Apache Spark erstellen](#)
- [Spark-Anwendungen mit Docker mithilfe von Amazon EMR 6.x ausführen](#)
- [Verwenden Sie den AWS Glue-Datenkatalog als Metastore für Spark SQL](#)
- [Konfigurieren von Spark](#)
- [Die Leistung von Spark optimieren](#)
- [Caching von Ergebnisfragmenten in Spark](#)
- [Den Nvidia-RAPIDS-Accelerator für Apache Spark verwenden](#)
- [Zugriff auf die Spark-Shell](#)
- [Verwenden Sie Amazon SageMaker Spark für maschinelles Lernen](#)
- [Eine Spark-Anwendung schreiben](#)
- [Verbessern der Spark Leistung mit Amazon S3](#)
- [Einen Spark-Schritt hinzufügen](#)

- [Anzeigen des Spark-Anwendungsverlaufs](#)
- [Zugreifen auf Web-Benutzerschnittstellen in Spark](#)
- [Verwenden des Amazon Kinesis Data Streams-Connectors für strukturiertes Streaming mit Spark](#)
- [Amazon-Redshift-Integration für Apache Spark mit Amazon EMR verwenden](#)
- [Spark-Versionsverlauf](#)

## Einen Cluster mit Apache Spark erstellen

Mit der folgenden Prozedur wird ein Cluster mit [Spark](#) erstellt, der mithilfe von Quick Options (Optionen für das schnelle Erstellen eines Clusters) in der Amazon-EMR-Konsole installiert wurde.

Sie können alternativ Erweiterte Optionen verwenden, um Ihre Cluster-Einrichtung weiter anzupassen oder Schritte zur programmgesteuerten Installation von Anwendungen und zur anschließenden Ausführung benutzerdefinierter Anwendungen zu senden. Bei diesen beiden erweiterten Optionen zur Clustererstellung können Sie AWS Glue als Spark-SQL-Metastore verwenden. Weitere Informationen finden Sie unter [Verwenden Sie den AWS Glue-Datenkatalog als Metastore für Spark SQL](#).

So starten Sie einen Cluster mit installiertem Spark

1. Öffnen Sie die Amazon EMR-Konsole unter <https://console.aws.amazon.com/emr>.
2. Wählen Sie Cluster erstellen aus, um Schnelle Optionen zu verwenden.
3. Geben Sie einen Clusternamen ein. Ihr Clustername darf die Zeichen <, >, \$, | oder ` (Backtick) nicht enthalten.
4. Wählen Sie unter Softwarekonfiguration die Option Version aus.
5. Wählen Sie für Anwendungen das Spark-Anwendungspaket.
6. Wählen Sie nach Bedarf weitere Optionen und anschließend Create cluster (Cluster erstellen) aus.

### Note

Weitere Informationen zum Konfigurieren von Spark beim Erstellen des Clusters finden Sie unter [Konfigurieren von Spark](#).

Um einen Cluster mit installiertem Spark zu starten, verwenden Sie AWS CLI

- Erstellen Sie den Cluster mit dem folgenden Befehl.

```
aws emr create-cluster --name "Spark cluster" --release-label emr-7.1.0 --
applications Name=Spark \
--ec2-attributes KeyName=myKey --instance-type m5.xlarge --instance-count 3 --use-
default-roles
```

### Note

Linux-Zeilenfortsetzungszeichen (\) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (^).

So starten Sie einen Cluster mit installiertem Spark über SDK für Java

Geben Sie Spark als Anwendung mit SupportedProductConfig an, die in RunJobFlowRequest verwendet wird.

- Das folgende Beispiel zeigt, wie Sie einen Cluster mit Spark mittels Java erstellen.

```
import com.amazonaws.AmazonClientException;
import com.amazonaws.auth.AWSCredentials;
import com.amazonaws.auth.AWSStaticCredentialsProvider;
import com.amazonaws.auth.profile.ProfileCredentialsProvider;
import com.amazonaws.services.elasticmapreduce.AmazonElasticMapReduce;
import com.amazonaws.services.elasticmapreduce.AmazonElasticMapReduceClientBuilder;
import com.amazonaws.services.elasticmapreduce.model.*;
import com.amazonaws.services.elasticmapreduce.util.StepFactory;

public class Main {

    public static void main(String[] args) {
        AWSCredentials credentials_profile = null;
        try {
            credentials_profile = new
ProfileCredentialsProvider("default").getCredentials();
        } catch (Exception e) {
```



```
        throw new AmazonClientException(
            "Cannot load credentials from .aws/
credentials file. " +
            "Make sure that the
credentials file exists and the profile name is specified within it.",
            e);
    }

    AmazonElasticMapReduce emr =
AmazonElasticMapReduceClientBuilder.standard()
        .withCredentials(new
AWSStaticCredentialsProvider(credentials_profile))
        .withRegion(Regions.US_WEST_1)
        .build();

    // create a step to enable debugging in the AWS Management Console
    StepFactory stepFactory = new StepFactory();
    StepConfig enableddebugging = new StepConfig()
        .withName("Enable debugging")
        .withActionOnFailure("TERMINATE_JOB_FLOW")

.withHadoopJarStep(stepFactory.newEnableDebuggingStep());

    Application spark = new Application().withName("Spark");

    RunJobFlowRequest request = new RunJobFlowRequest()
        .withName("Spark Cluster")
        .withReleaseLabel("emr-5.20.0")
        .withSteps(enableddebugging)
        .withApplications(spark)
        .withLogUri("s3://path/to/my/logs/")
        .withServiceRole("EMR_DefaultRole")
        .withJobFlowRole("EMR_EC2_DefaultRole")
        .withInstances(new JobFlowInstancesConfig()
            .withEc2SubnetId("subnet-12ab3c45")
            .withEc2KeyName("myEc2Key")
            .withInstanceCount(3)

.withKeepJobFlowAliveWhenNoSteps(true)
            .withMasterInstanceType("m4.large")

.withSlaveInstanceType("m4.large"));
    RunJobFlowResult result = emr.runJobFlow(request);
    System.out.println("The cluster ID is " + result.toString());
```

```
}  
}
```

## Spark-Anwendungen mit Docker mithilfe von Amazon EMR 6.x ausführen

Mit Amazon EMR 6.0.0 können Spark-Anwendungen Docker-Container für die Definition von Bibliotheksabhängigkeiten verwenden, anstatt Abhängigkeiten auf den einzelnen Amazon-EC2-Instances im Cluster zu installieren. Wenn Sie Spark mit Docker ausführen möchten, müssen Sie zunächst die Docker-Registrierung konfigurieren und beim Senden einer Spark-Anwendung zusätzliche Parameter definieren. Weitere Informationen finden Sie unter [Konfigurieren der Docker-Integration](#).

Wenn die Anwendung gesendet wird, ruft YARN Docker auf, um das angegebene Docker-Image abzurufen und die Spark-Anwendung in einem Docker-Container auszuführen. Auf diese Weise können Sie Abhängigkeiten einfach definieren und isolieren. Dadurch wird die Zeit reduziert, die für das Bootstrapping oder das Vorbereiten von Instances im Amazon-EMR-Cluster mit den für die Aufgabenausführung erforderlichen Bibliotheken notwendig ist.

### Überlegungen beim Ausführen von Spark mit Docker

Stellen Sie beim Ausführen von Spark mit Docker sicher, dass die folgenden Voraussetzungen erfüllt sind:

- Das `docker`-Paket und die CLI werden nur auf Kern- und Aufgabenknoten installiert.
- In Amazon EMR 6.1.0 und höher können Sie Docker alternativ mit den folgenden Befehlen auf einem Primärknoten installieren.
  - ```
sudo yum install -y docker  
sudo systemctl start docker
```
- Der `spark-submit`-Befehl sollte immer von einer primären-Instance auf dem Amazon-EMR-Cluster ausgeführt werden.
- Die Docker-Register, die zum Auflösen von Docker-Images verwendet werden, müssen mithilfe der Klassifikations-API mit dem `container-executor`-Klassifizierungsschlüssel definiert werden, um beim Starten des Clusters zusätzliche Parameter zu definieren:
  - `docker.trusted.registries`

- `docker.privileged-containers.registries`
- Wenn Sie eine Spark-Anwendung in einem Docker-Container ausführen möchten, sind die folgenden Konfigurationsoptionen erforderlich:
  - `YARN_CONTAINER_RUNTIME_TYPE=docker`
  - `YARN_CONTAINER_RUNTIME_DOCKER_IMAGE={DOCKER_IMAGE_NAME}`
- Beim Abrufen von Docker-Images mit Amazon ECR müssen Sie den Cluster so konfigurieren, dass er sich authentifiziert. Dazu müssen Sie die folgende Konfigurationsoption verwenden:
  - `YARN_CONTAINER_RUNTIME_DOCKER_CLIENT_CONFIG={DOCKER_CLIENT_CONFIG_PATH_ON_HDFS}`
- In Amazon EMR 6.1.0 und höher müssen Sie den aufgelisteten Befehl `YARN_CONTAINER_RUNTIME_DOCKER_CLIENT_CONFIG={DOCKER_CLIENT_CONFIG_PATH_ON_HDFS}` nicht verwenden, wenn das automatische ECR-Authentifizierungsfeature aktiviert ist.
- Bei jedem Docker-Image, das mit Spark verwendet wird, muss Java im Docker-Image installiert sein.

Weitere Informationen zu den Voraussetzungen finden Sie unter [Konfigurieren der Docker-Integration](#).

## Ein Docker-Image erstellen

Docker-Images werden mit einer Docker-Datei erstellt, in der die Pakete und die Konfiguration definiert werden, die in das Image aufgenommen werden sollen. Die folgenden beiden Beispiele verwenden Dockerfiles PySpark und SparkR.

### PySpark Docker-Datei

Docker-Images, die aus dieser Dockerfile erstellt wurden, enthalten Python 3 und das NumPy Python-Paket. Diese Docker-Datei verwendet Amazon Linux 2 und Amazon Corretto JDK 8.

```
FROM amazoncorretto:8

RUN yum -y update
RUN yum -y install yum-utils
RUN yum -y groupinstall development

RUN yum list python3*
RUN yum -y install python3 python3-dev python3-pip python3-virtualenv
```

```
RUN python -V
RUN python3 -V

ENV PYSPARK_DRIVER_PYTHON python3
ENV PYSPARK_PYTHON python3

RUN pip3 install --upgrade pip
RUN pip3 install numpy pandas

RUN python3 -c "import numpy as np"
```

## SparkR-Docker-Datei

Docker-Images, die aus dieser Docker-Datei erstellt werden, enthalten R und das CRAN-Paket „randomForest“. Diese Docker-Datei enthält Amazon Linux 2 und Amazon Corretto JDK 8.

```
FROM amazoncorretto:8

RUN java -version

RUN yum -y update
RUN amazon-linux-extras install R4

RUN yum -y install curl hostname

#setup R configs
RUN echo "r <- getOption('repos'); r['CRAN'] <- 'http://cran.us.r-project.org';
  options(repos = r);" > ~/.Rprofile

RUN Rscript -e "install.packages('randomForest')"
```

Weitere Informationen zur Docker-Dateisyntax finden Sie in der [Docker-Datei-Referenzdokumentation](#).

## Verwenden von Docker-Images von Amazon ECR

Amazon Elastic Container Registry (Amazon ECR) ist eine vollständig verwaltete Docker-Container-Registrierung, die das Speichern, Verwalten und Bereitstellen von Docker-Container-Images erleichtert. Wenn Sie Amazon ECR verwenden, muss der Cluster so konfiguriert sein, dass er Ihrer ECR-Instance vertraut, und Sie müssen die Authentifizierung konfigurieren, damit der Cluster Docker-

Images aus Amazon ECR verwenden kann. Weitere Informationen finden Sie unter [Konfigurieren von YARN für den Zugriff auf Amazon ECR](#).

Um sicherzustellen, dass Amazon-EMR-Hosts auf die in Amazon ECR gespeicherten Images zugreifen können, muss der Cluster über die Berechtigungen der AmazonEC2ContainerRegistryReadOnly-Richtlinie verfügen, die dem Instance-Profil zugeordnet ist. Weitere Informationen finden Sie unter [AmazonEC2ContainerRegistryReadOnly-Richtlinie](#).

In diesem Beispiel muss der Cluster mit der folgenden zusätzlichen Konfiguration erstellt werden, um sicherzustellen, dass die Amazon-ECR-Registrierung vertrauenswürdig ist. Ersetzen Sie den Endpunkt *123456789123.dkr.ecr.us-east-1.amazonaws.com* durch Ihren Amazon-ECR-Endpunkt.

```
[
  {
    "Classification": "container-executor",
    "Configurations": [
      {
        "Classification": "docker",
        "Properties": {
          "docker.privileged-containers.registries":
"local,centos,123456789123.dkr.ecr.us-east-1.amazonaws.com",
          "docker.trusted.registries": "local,centos,123456789123.dkr.ecr.us-
east-1.amazonaws.com"
        }
      }
    ],
    "Properties": {}
  }
]
```

## Verwendung PySpark mit Amazon ECR

Im folgenden Beispiel wird das PySpark Dockerfile verwendet, das markiert und in Amazon ECR hochgeladen wird. Nachdem Sie das Dockerfile hochgeladen haben, können Sie den PySpark Job ausführen und auf das Docker-Image von Amazon ECR verweisen.

Nachdem Sie den Cluster gestartet haben, verwenden Sie SSH, um eine Verbindung zu einem Kernknoten herzustellen, und führen Sie die folgenden Befehle aus, um das lokale Docker-Image aus dem Dockerfile-Beispiel zu erstellen. PySpark

Erstellen Sie zunächst ein Verzeichnis und eine Docker-Datei.

```
mkdir pyspark
vi pyspark/Dockerfile
```

Fügen Sie den Inhalt der PySpark Dockerfile ein und führen Sie die folgenden Befehle aus, um ein Docker-Image zu erstellen.

```
sudo docker build -t local/pyspark-example pyspark/
```

Erstellen Sie das `emr-docker-examples`-ECR-Repository für die Beispiele.

```
aws ecr create-repository --repository-name emr-docker-examples
```

Markieren Sie das vor Ort erstellte Image und laden Sie es in ECR hoch, indem Sie `123456789123.dkr.ecr.us-east-1.amazonaws.com` durch Ihren ECR-Endpunkt ersetzen.

```
sudo docker tag local/pyspark-example 123456789123.dkr.ecr.us-east-1.amazonaws.com/emr-docker-examples:pyspark-example
sudo docker push 123456789123.dkr.ecr.us-east-1.amazonaws.com/emr-docker-examples:pyspark-example
```

Verwenden Sie SSH, um eine Verbindung mit dem Primärknoten herzustellen und ein Python-Skript mit dem Dateinamen „`main.py`“ vorzubereiten. Fügen Sie den folgenden Inhalt in die `main.py`-Datei ein und speichern Sie sie.

```
from pyspark.sql import SparkSession
spark = SparkSession.builder.appName("docker-numpy").getOrCreate()
sc = spark.sparkContext

import numpy as np
a = np.arange(15).reshape(3, 5)
print(a)
```

Verweisen Sie unter Amazon EMR 6.0.0 zum Übermitteln des Auftrags auf den Namen des Docker-Images. Definieren Sie die zusätzlichen Konfigurationsparameter, um sicherzustellen, dass die Aufgabenausführung Docker als Laufzeit verwendet. Bei Verwendung von Amazon ECR muss `YARN_CONTAINER_RUNTIME_DOCKER_CLIENT_CONFIG` sich auf die `config.json`-Datei

beziehen, in der die Anmeldeinformationen enthalten sind, die zur Authentifizierung bei Amazon ECR verwendet werden.

```

DOCKER_IMAGE_NAME=123456789123.dkr.ecr.us-east-1.amazonaws.com/emr-docker-
examples:pyspark-example
DOCKER_CLIENT_CONFIG=hdfs:///user/hadoop/config.json
spark-submit --master yarn \
--deploy-mode cluster \
--conf spark.executorEnv.YARN_CONTAINER_RUNTIME_TYPE=docker \
--conf spark.executorEnv.YARN_CONTAINER_RUNTIME_DOCKER_IMAGE=$DOCKER_IMAGE_NAME \
--conf spark.executorEnv.YARN_CONTAINER_RUNTIME_DOCKER_CLIENT_CONFIG=
$DOCKER_CLIENT_CONFIG \
--conf spark.yarn.appMasterEnv.YARN_CONTAINER_RUNTIME_TYPE=docker \
--conf spark.yarn.appMasterEnv.YARN_CONTAINER_RUNTIME_DOCKER_IMAGE=$DOCKER_IMAGE_NAME \
--conf spark.yarn.appMasterEnv.YARN_CONTAINER_RUNTIME_DOCKER_CLIENT_CONFIG=
$DOCKER_CLIENT_CONFIG \
--num-executors 2 \
main.py -v

```

Verweisen Sie unter Amazon EMR 6.1.0 oder höher zum Übermitteln des Auftrags auf den Namen des Docker-Images. Wenn die automatische ECR-Authentifizierung aktiviert ist, führen Sie den folgenden Befehl aus.

```

DOCKER_IMAGE_NAME=123456789123.dkr.ecr.us-east-1.amazonaws.com/emr-docker-
examples:pyspark-example
spark-submit --master yarn \
--deploy-mode cluster \
--conf spark.executorEnv.YARN_CONTAINER_RUNTIME_TYPE=docker \
--conf spark.executorEnv.YARN_CONTAINER_RUNTIME_DOCKER_IMAGE=$DOCKER_IMAGE_NAME \
--conf spark.yarn.appMasterEnv.YARN_CONTAINER_RUNTIME_TYPE=docker \
--conf spark.yarn.appMasterEnv.YARN_CONTAINER_RUNTIME_DOCKER_IMAGE=$DOCKER_IMAGE_NAME \
--num-executors 2 \
main.py -v

```

Wenn der Job abgeschlossen ist, notieren Sie sich die YARN-Anwendungs-ID und verwenden Sie den folgenden Befehl, um die Ausgabe des Jobs abzurufen. PySpark

```

yarn logs --applicationId application_id | grep -C2 '\[ \['
LogLength:55
LogContents:
[[ 0  1  2  3  4]
 [ 5  6  7  8  9]

```

```
[10 11 12 13 14]]
```

## SparkR mit Amazon ECR verwenden

Im folgenden Beispiel wird die SparkR-Docker-Datei verwendet, die markiert und in ECR hochgeladen wird. Sobald die Docker-Datei hochgeladen wurde, können Sie die SparkR-Aufgabe ausführen und auf das Docker-Image von Amazon ECR verweisen.

Nachdem Sie den Cluster gestartet haben, verwenden Sie SSH, um eine Verbindung mit einem Core-Knoten herzustellen, und führen Sie die folgenden Befehle aus, um das lokale Docker-Image aus dem Beispiel für die SparkR-Docker-Datei zu erstellen.

Erstellen Sie zunächst ein Verzeichnis und die Docker-Datei.

```
mkdir sparkr
vi sparkr/Dockerfile
```

Fügen Sie den Inhalt der SparkR-Docker-Datei ein, und führen Sie die folgenden Befehle aus, um ein Docker-Image zu erstellen.

```
sudo docker build -t local/sparkr-example sparkr/
```

Markieren Sie das vor Ort erstellte Image und laden Sie es in Amazon ECR hoch, indem Sie [123456789123.dkr.ecr.us-east-1.amazonaws.com](https://123456789123.dkr.ecr.us-east-1.amazonaws.com) durch Ihren Amazon-ECR-Endpunkt ersetzen.

```
sudo docker tag local/sparkr-example 123456789123.dkr.ecr.us-east-1.amazonaws.com/emr-
docker-examples:sparkr-example
sudo docker push 123456789123.dkr.ecr.us-east-1.amazonaws.com/emr-docker-
examples:sparkr-example
```

Verwenden Sie SSH, um eine Verbindung zum Primärknoten herzustellen und ein R-Skript mit dem Namen `sparkR.R` vorzubereiten. Fügen Sie den folgenden Inhalt in die `sparkR.R`-Datei ein:

```
library(SparkR)
sparkR.session(appName = "R with Spark example", sparkConfig =
  list(spark.some.config.option = "some-value"))

sqlContext <- sparkRSQL.init(spark.sparkContext)
```



```
library(randomForest)
# check release notes of randomForest
rfNews()

sparkR.session.stop()
```

Verweisen Sie unter Amazon EMR 6.0.0 zum Übermitteln des Auftrags auf den Namen des Docker-Images. Definieren Sie die zusätzlichen Konfigurationsparameter, um sicherzustellen, dass die Aufgabenausführung Docker als Laufzeit verwendet. Bei Verwendung von Amazon ECR muss `YARN_CONTAINER_RUNTIME_DOCKER_CLIENT_CONFIG` sich auf die `config.json`-Datei beziehen, in der die Anmeldeinformationen enthalten sind, die zur Authentifizierung bei ECR verwendet werden.

```
DOCKER_IMAGE_NAME=123456789123.dkr.ecr.us-east-1.amazonaws.com/emr-docker-
examples:sparkr-example
DOCKER_CLIENT_CONFIG=hdfs:///user/hadoop/config.json
spark-submit --master yarn \
--deploy-mode cluster \
--conf spark.executorEnv.YARN_CONTAINER_RUNTIME_TYPE=docker \
--conf spark.executorEnv.YARN_CONTAINER_RUNTIME_DOCKER_IMAGE=$DOCKER_IMAGE_NAME \
--conf spark.executorEnv.YARN_CONTAINER_RUNTIME_DOCKER_CLIENT_CONFIG=
$DOCKER_CLIENT_CONFIG \
--conf spark.yarn.appMasterEnv.YARN_CONTAINER_RUNTIME_TYPE=docker \
--conf spark.yarn.appMasterEnv.YARN_CONTAINER_RUNTIME_DOCKER_IMAGE=$DOCKER_IMAGE_NAME \
--conf spark.yarn.appMasterEnv.YARN_CONTAINER_RUNTIME_DOCKER_CLIENT_CONFIG=
$DOCKER_CLIENT_CONFIG \
sparkR.R
```

Verweisen Sie unter Amazon EMR 6.1.0 oder höher zum Übermitteln des Auftrags auf den Namen des Docker-Images. Wenn die automatische ECR-Authentifizierung aktiviert ist, führen Sie den folgenden Befehl aus.

```
DOCKER_IMAGE_NAME=123456789123.dkr.ecr.us-east-1.amazonaws.com/emr-docker-
examples:sparkr-example
spark-submit --master yarn \
--deploy-mode cluster \
--conf spark.executorEnv.YARN_CONTAINER_RUNTIME_TYPE=docker \
--conf spark.executorEnv.YARN_CONTAINER_RUNTIME_DOCKER_IMAGE=$DOCKER_IMAGE_NAME \
--conf spark.yarn.appMasterEnv.YARN_CONTAINER_RUNTIME_TYPE=docker \
--conf spark.yarn.appMasterEnv.YARN_CONTAINER_RUNTIME_DOCKER_IMAGE=$DOCKER_IMAGE_NAME \
sparkR.R
```

Wenn die Aufgabe abgeschlossen ist, notieren Sie sich die YARN-Anwendungs-ID, und verwenden Sie den folgenden Befehl, um die Ausgabe der SparkR-Aufgabe abzurufen. Dieses Beispiel enthält Tests, um sicherzustellen, dass die randomForest-Bibliothek, die installierte Version und die Versionshinweise verfügbar sind.

```
yarn logs --applicationId application_id | grep -B4 -A10 "Type rfNews"
randomForest 4.6-14
Type rfNews() to see new features/changes/bug fixes.
Wishlist (formerly TODO):

* Implement the new scheme of handling classwt in classification.

* Use more compact storage of proximity matrix.

* Allow case weights by using the weights in sampling?

=====
Changes in 4.6-14:
```

## Verwenden Sie den AWS Glue-Datenkatalog als Metastore für Spark SQL

Mit Amazon EMR Version 5.8.0 oder höher können Sie Spark SQL so konfigurieren, dass der AWS Glue-Datenkatalog als Metastore verwendet wird. Wir empfehlen diese Konfiguration, wenn Sie einen persistenten Metastore oder einen gemeinsamen Metastore für verschiedene Cluster, Services, Anwendungen oder AWS -Konten benötigen.

AWS Glue ist ein vollständig verwalteter ETL-Service (Extrahieren, Transformieren und Laden), mit dem Sie Ihre Daten einfach und kostengünstig kategorisieren, bereinigen, anreichern und zuverlässig zwischen verschiedenen Datenspeichern verschieben können. Der AWS Glue Data Catalog bietet ein einheitliches Metadaten-Repository für eine Vielzahl von Datenquellen und Datenformaten, das sich in Amazon EMR sowie Amazon RDS, Amazon Redshift, Redshift Spectrum, Athena und jede mit dem Apache Hive Metastore kompatible Anwendung integrieren lässt. AWS Glue-Crawler können automatisch Schemas aus Quelldaten in Amazon S3 ableiten und die zugehörigen Metadaten im Datenkatalog speichern. Weitere Informationen zum Datenkatalog finden Sie unter [Füllen des AWS Glue-Datenkatalogs im AWS Glue-Entwicklerhandbuch](#).

Für AWS Glue fallen separate Gebühren an. Es gibt eine monatliche Gebühr für das Speichern und Zugreifen auf die Metadaten im Datenkatalog, einen Stundensatz, der pro Minute für AWS Glue

ETL-Jobs und Crawler-Laufzeit abgerechnet wird, und einen Stundensatz, der pro Minute für jeden bereitgestellten Entwicklungsendpunkt abgerechnet wird. Der Datenkatalog ermöglicht Ihnen die gebührenfreie Speicherung von bis zu einer Million Objekten. Wenn Sie mehr als eine Million Objekte speichern, fällt eine Gebühr von 1 USD für jeweils 100.000 zusätzliche Objekte an. Im Data Catalog ist ein Objekt eine Tabelle, eine Partition oder eine Datenbank. Weitere Informationen finden Sie unter [Glue-Preise](#).

#### Important

Wenn Sie vor dem 14. August 2017 Tabellen mit Amazon Athena oder Amazon Redshift Spectrum erstellt haben, werden Datenbanken und Tabellen in einem von Athena verwalteten Katalog gespeichert, der vom Glue-Datenkatalog getrennt ist. AWS Um Amazon EMR mit diesen Tabellen zu integrieren, müssen Sie ein Upgrade auf den AWS Glue Data Catalog durchführen. Weitere Informationen finden Sie unter [Upgrade auf den AWS Glue-Datenkatalog](#) im Amazon Athena Athena-Benutzerhandbuch.

## AWS Glue Data Catalog als Metastore angeben

Sie können den AWS Glue-Datenkatalog mithilfe der Amazon EMR-API AWS Management Console AWS CLI, oder als Metastore angeben. Wenn Sie die CLI oder API verwenden, verwenden Sie die Konfigurationsklassifizierung für Spark, um den Datenkatalog anzugeben. Darüber hinaus können Sie mit Amazon EMR 5.16.0 und höher die Konfigurationsklassifizierung verwenden, um einen Datenkatalog in einem anderen Format anzugeben. AWS-KontoWenn Sie die Konsole verwenden, können Sie den Data Catalog mit den erweiterten Optionen oder den Schnelloptionen angeben.

#### Note

Die Option zur Verwendung von AWS Glue Data Catalog ist auch mit Zeppelin verfügbar, da Zeppelin mit Spark-SQL-Komponenten installiert ist.

### New console

So geben Sie AWS Glue Data Catalog mit der neuen Konsole als Spark-Metastore an

1. Melden Sie sich bei der AWS Management Console an und öffnen Sie die Amazon EMR-Konsole unter <https://console.aws.amazon.com/emr>.

2. Wählen Sie im linken Navigationsbereich unter Amazon EMR in EC2 die Option Cluster und dann Create cluster (Cluster erstellen) aus.
3. Wählen Sie unter Anwendungspaket die Option Spark oder Benutzerdefiniert aus. Wenn Sie Ihren Cluster anpassen, stellen Sie sicher, dass Sie Zeppelin oder Spark als eine Ihrer Anwendungen auswählen.
4. Aktivieren Sie unter Einstellungen für den AWS Glue Data Catalog das Kontrollkästchen Für Spark-Tabellenmetadaten verwenden.
5. Wählen Sie alle anderen Optionen aus, die für Ihren Cluster gelten.
6. Um Ihren Cluster jetzt zu starten, wählen Sie Cluster erstellen aus.

## Old console

Um AWS Glue Data Catalog mit der alten Konsole als Spark-Metastore zu spezifizieren

1. Navigieren Sie zur neuen Amazon-EMR-Konsole und wählen Sie in der Seitennavigation die Option Zur alten Konsole wechseln aus. Weitere Informationen darüber, was Sie erwartet, wenn Sie zur alten Konsole wechseln, finden Sie unter [Verwenden der alten Konsole](#).
2. Wählen Sie Create Cluster (Cluster erstellen) und Go to advanced options (Zu erweiterten Optionen) aus.
3. Wählen Sie für Release die Option emr-5.8.0 oder höher aus.
4. Wählen Sie unter Release eine der Optionen Spark oder Zeppelin aus.
5. Wählen Sie unter Einstellungen für AWS Glue Data Catalog die Option Für Spark-Tabellenmetadaten verwenden.
6. Wählen Sie weitere Optionen für Ihren Cluster nach Bedarf aus, wählen Sie Next (Weiter) aus und konfigurieren Sie dann weitere Cluster-Optionen gemäß Ihrer Anwendung.

## AWS CLI

Um den AWS Glue-Datenkatalog als Spark-Metastore zu spezifizieren mit dem AWS CLI

Weitere Informationen zur Angabe einer Konfigurationsklassifizierung mithilfe der AWS CLI Amazon EMR-API finden Sie unter [Anwendungen konfigurieren](#).

- Geben Sie den Wert für `hive.metastore.client.factory.class` mit der Klassifizierung `spark-hive-site` an, wie in dem folgenden Beispiel gezeigt:

```
[
  {
    "Classification": "spark-hive-site",
    "Properties": {
      "hive.metastore.client.factory.class":
"com.amazonaws.glue.catalog.metastore.AWSGlueDataCatalogHiveClientFactory"
    }
  }
]
```

Um einen Datenkatalog in einem anderen AWS Konto anzugeben, fügen Sie die `hive.metastore.glue.catalogid` Eigenschaft hinzu, wie im folgenden Beispiel gezeigt. Ersetzen Sie *acct-id* durch das AWS -Konto des Data Catalog.

```
[
  {
    "Classification": "spark-hive-site",
    "Properties": {
      "hive.metastore.client.factory.class":
"com.amazonaws.glue.catalog.metastore.AWSGlueDataCatalogHiveClientFactory",
      "hive.metastore.glue.catalogid": "acct-id"
    }
  }
]
```

## IAM-Berechtigungen

Das EC2-Instanzprofil für einen Cluster muss über IAM-Berechtigungen für AWS Glue-Aktionen verfügen. Wenn Sie die Verschlüsselung für AWS Glue Data Catalog-Objekte aktivieren, muss die Rolle außerdem berechtigt sein, die für die Verschlüsselung AWS KMS key verwendeten Daten zu verschlüsseln, zu entschlüsseln und zu generieren.

### Berechtigungen für AWS Glue-Aktionen

Wenn Sie das standardmäßige EC2-Instance-Profil für Amazon EMR verwenden, ist keine Aktion erforderlich. Die `AmazonElasticMapReduceforEC2Role` verwaltete Richtlinie, die an die angehängt ist, `EMR_EC2_DefaultRole` ermöglicht alle erforderlichen AWS Glue-Aktionen. Wenn

Sie jedoch ein benutzerdefiniertes EC2-Instanzprofil und Berechtigungen angeben, müssen Sie die entsprechenden AWS Glue-Aktionen konfigurieren. Verwenden Sie hierzu die verwaltete Richtlinie `AmazonElasticMapReduceforEC2Role` als Ausgangspunkt. Weitere Informationen finden Sie unter [Servicerolle für Cluster-EC2-Instances \(EC2-Instance-Profil\)](#) im Verwaltungshandbuch für Amazon EMR.

## Berechtigungen zum Verschlüsseln und Entschlüsseln von AWS Glue Data Catalog

Ihr Instance-Profil benötigt die Erlaubnis, Daten mithilfe Ihres Schlüssels zu verschlüsseln und zu entschlüsseln. Sie müssen diese Berechtigungen nicht konfigurieren, wenn die beiden folgenden Aussagen zutreffen:

- Sie aktivieren die Verschlüsselung für AWS Glue Data Catalog-Objekte mithilfe von verwalteten Schlüsseln für AWS Glue.
- Sie verwenden einen Cluster, der sich im selben Verzeichnis befindet AWS-Konto wie der AWS Glue-Datenkatalog.

Andernfalls müssen Sie die folgende Anweisung der Berechtigungsrichtlinie hinzufügen, die mit Ihrem EC2-Instance-Profil verknüpft ist.

```
[
  {
    "Version": "2012-10-17",
    "Statement": [
      {
        "Effect": "Allow",
        "Action": [
          "kms:Decrypt",
          "kms:Encrypt",
          "kms:GenerateDataKey"
        ],
        "Resource": "arn:aws:kms:region:acct-
id:key/12345678-1234-1234-1234-123456789012"
      }
    ]
  }
]
```

Weitere Informationen zur Verschlüsselung von AWS Glue Data Catalog finden Sie unter [Verschlüsselung Ihres Datenkatalogs](#) im AWS Glue Developer Guide.

## Ressourcenbasierte Berechtigungen

Wenn Sie AWS Glue in Verbindung mit Hive, Spark oder Presto in Amazon EMR verwenden, unterstützt AWS Glue ressourcenbasierte Richtlinien zur Steuerung des Zugriffs auf Datenkatalogressourcen. Zu diesen Ressourcen gehören Datenbanken, Tabellen, Verbindungen und benutzerdefinierte Funktionen. Weitere Informationen finden Sie unter [Verwenden von ressourcenbasierten Richtlinien für AWS Glue](#) im AWS -Glue-Entwicklerhandbuch.

Wenn Sie ressourcenbasierte Richtlinien verwenden, um den Zugriff auf AWS Glue von Amazon EMR aus zu beschränken, muss der Principal, den Sie in der Berechtigungsrichtlinie angeben, der Rollen-ARN sein, der dem EC2-Instance-Profil zugeordnet ist, das bei der Erstellung eines Clusters angegeben wird. Beispielsweise können Sie für eine ressourcenbasierte Richtlinie, die an einen Katalog angehängt ist, den Rollen-ARN für die Standarddienstrolle für Cluster-EC2-Instances, `EMR_EC2_`, `DefaultRole` als angeben, indem Sie das im folgenden Beispiel gezeigte Principal Format verwenden:

```
arn:aws:iam::acct-id:role/EMR_EC2_DefaultRole
```

Die *Konto-ID* kann sich von der AWS Glue-Konto-ID unterscheiden. Dies ermöglicht den Zugriff von EMR-Clustern in verschiedenen Konten aus. Sie können mehrere Principals angeben, von denen jeder aus einem anderen Konto stammt.

## Überlegungen zur Verwendung von AWS Glue Data Catalog

Beachten Sie die folgenden Punkte, wenn Sie AWS Glue Data Catalog als Metastore mit Spark verwenden:

- Das Vorhandensein einer Standarddatenbank ohne Speicherort-URI sorgt für Fehler beim Erstellen einer Tabelle. Um dieses Problem zu umgehen, verwenden Sie die LOCATION-Klausel, um den Bucket-Speicherort anzugeben, z. B. `s3://EXAMPLE-DOC-BUCKET`, wenn Sie CREATE TABLE verwenden. Alternativ können Sie auch Tabellen in einer anderen Datenbank als der Standarddatenbank erstellen.
- Das Umbenennen von Tabellen in AWS Glue wird nicht unterstützt.
- Wenn Sie eine Hive-Tabelle erstellen, ohne eine LOCATION anzugeben, werden die Tabellendaten an dem in der Eigenschaft `hive.metastore.warehouse.dir` angegebenen Ort gespeichert. Standardmäßig ist dies ein Speicherort in HDFS. Wenn ein anderer Cluster auf die Tabelle zugreifen muss, schlägt er fehl, sofern er nicht über ausreichende Berechtigungen für den Cluster verfügt, der die Tabelle erstellt hat. Da HDFS-Speicher vorübergehend ist, gehen außerdem

die Tabellendaten verloren, wenn der Cluster beendet wird, und die Tabelle muss neu erstellt werden. Wir empfehlen, dass Sie LOCATION in Amazon S3 angeben, wenn Sie eine Hive-Tabelle mit AWS Glue erstellen. Alternativ können Sie die hive-site-Konfigurationsklassifizierung verwenden, um einen Speicherort in Amazon S3 für hive.metastore.warehouse.dir anzugeben, der für alle Hive-Tabellen gilt. Wenn eine Tabelle an einem HDFS-Speicherort erstellt wurde und der Cluster, der sie erstellt hat, noch läuft, können Sie den Tabellenspeicherort von AWS Glue aus auf Amazon S3 aktualisieren. Weitere Informationen finden Sie unter [Arbeiten mit Tabellen auf der AWS Glue-Konsole](#) im AWS Glue-Entwicklerhandbuch.

- Partitionswerte, die Anführungszeichen und Apostrophe enthalten, werden nicht unterstützt, zum Beispiel PARTITION (owner="Doe 's").
- [Spaltenstatistiken](#) werden für emr-5.31.0 und höher unterstützt.
- Die Verwendung der [Hive-Autorisierung](#) wird nicht unterstützt. Als Alternative sollten Sie die Verwendung [ressourcenbasierter Richtlinien für AWS Glue](#) in Betracht ziehen. Weitere Informationen finden Sie unter [Verwenden von ressourcenbasierten Richtlinien für Amazon EMR Access to AWS Glue Data Catalog](#).

## Konfigurieren von Spark

Sie können [Spark in Amazon EMR](#) mithilfe von Konfigurationsklassifizierungen konfigurieren. Weitere Informationen über Konfigurationsklassifizierungen finden Sie unter [Anwendungen konfigurieren](#).

Zu den Konfigurationsklassifizierungen für Spark in Amazon EMR gehören die folgenden:

- **spark** – Setzt die maximizeResourceAllocation-Eigenschaft auf „wahr“ oder „falsch“. Wenn „wahr“, konfiguriert Amazon EMR die spark-defaults-Eigenschaften automatisch auf der Grundlage der Cluster-Hardwarekonfiguration. Weitere Informationen finden Sie unter [Verwenden von maximizeResourceAllocation](#).
- **spark-defaults** – Setzt Werte in der spark-defaults.conf-Datei. Weitere Informationen finden Sie unter [Spark-Konfiguration](#) in der Spark-Dokumentation.
- **spark-env** – Setzt Werte in der spark-env.sh-Datei. Weitere Informationen finden Sie unter [Umgebungsvariablen](#) in der Spark-Dokumentation.
- **spark-hive-site** – Legt Werte in der hive-site.xml für Spark fest.
- **spark-log4j** – (Amazon-EMR-Versionen 6.7.x und niedriger) Legt Werte in der log4j.properties-Datei fest. Weitere Informationen finden Sie in der [log4j.properties.template](#)-Datei auf Github.



- **spark-log4j2** – (Amazon-EMR-Version 6.8.0 und höher) Legt Werte in der `log4j2.properties`-Datei fest. Weitere Informationen finden Sie in der [log4j2.properties.template](#)-Datei auf Github.
- **spark-metrics** – Setzt Werte in der `metrics.properties`-Datei. Einstellungen und weitere Informationen finden Sie in der [metrics.properties.template](#)-Datei auf GitHub und unter [Metrics](#) in der Spark-Dokumentation.

### Note

Wenn Sie Spark-Workloads von einer anderen Plattform zu Amazon EMR migrieren, empfehlen wir Ihnen, Ihre Workloads mit den [Durch Amazon EMR festgelegte Spark-StandardEinstellungen](#) zu testen, bevor Sie benutzerdefinierte Konfigurationen hinzufügen. Die meisten Kunden sehen mit unseren StandardEinstellungen eine verbesserte Leistung.

## Themen


- [Durch Amazon EMR festgelegte Spark-StandardEinstellungen](#)
- [Konfiguration der Spark-Garbage-Collection auf Amazon EMR 6.1.0](#)
- [Verwenden von maximizeResourceAllocation](#)
- [Verhalten der Außerbetriebnahme von Knoten konfigurieren](#)
- [ThriftServer Spark-Umgebungsvariable](#)
- [Ändern von Spark-StandardEinstellungen](#)
- [Migration von Apache Log4j 1.x zu Log4j 2.x](#)


## Durch Amazon EMR festgelegte Spark-StandardEinstellungen

Die folgende Tabelle zeigt, wie Amazon EMR Standardwerte in `spark-defaults` setzt, die sich auf Anwendungen auswirken.

### Durch Amazon EMR festgelegte Spark-StandardEinstellungen

| Einstellung                        | Beschreibung                                      | Standardwert                                         |
|------------------------------------|---------------------------------------------------|------------------------------------------------------|
| <code>spark.executor.memory</code> | Die pro Executor-Prozess zu verwendende Größe des | Diese Einstellung wird durch die Core- und Aufgaben- |

| Einstellung                                                            | Beschreibung                                                                                                                                                                               | Standardwert                                                                                                                                                                                                                                                                                      |
|------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                                                                        | Speichers. Zum Beispiel: 1g, 2g.                                                                                                                                                           | Instance-Typen im Cluster bestimmt.                                                                                                                                                                                                                                                               |
| <code>spark.executor.cores</code>                                      | Die für jeden Executor zu verwendende Anzahl der Kerne.                                                                                                                                    | Diese Einstellung wird durch die Core- und Aufgaben-Instance-Typen im Cluster bestimmt.                                                                                                                                                                                                           |
| <code>spark.dynamicAllocation.enabled</code>                           | Wenn „wahr“, verwenden Sie die dynamische Ressourcenzuteilung, um die Anzahl der bei einer Anwendung registrierten Executors basierend auf dem Workload nach oben oder unten zu skalieren. | <code>true</code> (mit Amazon EMR 4.4.0 und höher)                                                                                                                                                                                                                                                |
|                                                                        |                                                                                                                                                                                            | <div data-bbox="1068 688 1507 1003" style="border: 1px solid #add8e6; border-radius: 10px; padding: 10px;"> <p> <b>Note</b><br/>Spark Shuffle Service wird automatisch von Amazon EMR konfiguriert.</p> </div> |
| <code>spark.sql.hive.advancedPartitionPredicatePushdown.enabled</code> | Wenn der Wert „wahr“ ist, ist der erweiterte Pushdown von Partitionsprädikaten in den Hive-Metastore aktiviert.                                                                            | <code>true</code>                                                                                                                                                                                                                                                                                 |

| Einstellung                                                              | Beschreibung                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | Standardwert             |
|--------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------|
| <code>spark.sql.hive.stringLikePartitionPredicatePushdown.enabled</code> | <p>Verschiebt und filtert <code>startsWith</code>, <code>contains</code> und <code>endsWith</code> nach unten in den Hive-Metastore.</p> <div data-bbox="591 445 1029 1241" style="border: 1px solid #add8e6; border-radius: 10px; padding: 10px; margin-top: 10px;"> <p> <b>Note</b></p> <p>Glue unterstützt das Prädikat-Pushdown für <code>startsWith</code>, <code>contains</code> oder <code>endsWith</code> nicht. Wenn Sie Glue Metastore verwenden und aufgrund des Prädikat-Pushdowns für diese Funktionen auf Fehler stoßen, setzen Sie diese Konfiguration auf <code>false</code>.</p> </div> | <p><code>true</code></p> |

## Konfiguration der Spark-Garbage-Collection auf Amazon EMR 6.1.0

Das Einrichten von benutzerdefinierten Garbage-Collection-Konfigurationen mit `spark.driver.extraJavaOptions` und `spark.executor.extraJavaOptions` führt dazu, dass der Treiber- oder Ausführer-Start mit Amazon EMR 6.1 fehlschlägt, weil eine Konfiguration der Garbage-Collection-Konfiguration mit Amazon EMR 6.1.0 in Konflikt steht. Für Amazon EMR 6.1.0 wird die Standard-Garbage-Collection über `spark.driver.defaultJavaOptions` und `spark.executor.defaultJavaOptions` festgelegt. Diese Konfiguration gilt nur für Amazon EMR 6.1.0. JVM-Optionen, die nichts mit der Garbage-Collection zu tun haben, wie z. B. die für die Konfiguration von Protokollierung (`-verbose: class`), können dennoch durch `extraJavaOptions` festgelegt werden. Weitere Informationen finden Sie unter [Spark-Anwendungseigenschaften](#).

## Verwenden von `maximizeResourceAllocation`

Um Ihre Executors so zu konfigurieren, dass sie auf jedem Knoten in einem Cluster die maximal möglichen Ressourcen nutzen, setzen Sie in Ihrer `spark`-Konfigurationsklassifizierung `maximizeResourceAllocation` auf `true`. `maximizeResourceAllocation` ist spezifisch für Amazon EMR. Wenn Sie `maximizeResourceAllocation` aktivieren, berechnet Amazon EMR die maximal verfügbaren Rechen- und Speicherressourcen für einen Executor auf einer Instance in der Core-Instance-Gruppe. Anschließend werden die entsprechenden `spark-defaults`-Einstellungen auf der Grundlage der berechneten Höchstwerte festgelegt.

Amazon EMR berechnet die maximalen Rechen- und Speicherressourcen, die für einen Executor verfügbar sind, basierend auf einem Instance-Typ aus der Core-Instance-Flotte. Da jede Instance-Flotte unterschiedliche Instance-Typen und Größen innerhalb einer Flotte haben kann, ist die von Amazon EMR verwendete Executor-Konfiguration möglicherweise nicht die beste für Ihre Cluster. Daher empfehlen wir nicht, die Standardeinstellungen zu verwenden, wenn Sie die maximale Ressourcenzuweisung verwenden. Konfigurieren Sie benutzerdefinierte Einstellungen für Ihre Instance-Flottencluster.

### Note

Sie sollten die `maximizeResourceAllocation`-Option nicht für Cluster mit anderen verteilten Anwendungen wie HBase verwenden. Amazon EMR verwendet benutzerdefinierte YARN-Konfigurationen für verteilte Anwendungen, die zu Konflikten mit `maximizeResourceAllocation` führen und zum Scheitern von Spark-Anwendungen führen können.

Im Folgenden finden Sie ein Beispiel für eine Spark-Konfigurationsklassifizierung mit der `maximizeResourceAllocation`-Einstellung auf `true`.

```
[
  {
    "Classification": "spark",
    "Properties": {
      "maximizeResourceAllocation": "true"
    }
  }
]
```

Einstellungen, die in **spark-defaults** konfiguriert sind, wenn **maximizeResourceAllocation** aktiviert ist

| Einstellung               | Beschreibung                                                                                                                                                                | Wert                                                                                                                                                                                                                                                                                                                                                             |
|---------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| spark.default.parallelism | Standardanzahl von Partitionen in RDDs, die von Transformationen wie Join, und Parallelize zurückgegeben werden reduceByKey, wenn sie nicht vom Benutzer festgelegt wurden. | Die zweifache Anzahl von CPU-Kernen für YARN-Container.                                                                                                                                                                                                                                                                                                          |
| spark.driver.memory       | Speichermenge, die für den Treiberprozess verwendet werden soll, d. h. wo er initialisiert wird. SparkContext (zum Beispiel 1 g, 2 g).                                      | Die Einstellung wird auf der Grundlage der Instance-Typen im Cluster konfiguriert. Da die Spark-Treiberanwendung jedoch entweder in der Primär-Instance oder in einer der Core-Instances ausgeführt werden kann (z. B. im YARN-Client- und -Cluster-Modus), wird sie auf der Grundlage des kleineren Instance-Typs in diesen beiden Instance-Gruppen festgelegt. |
| spark.executor.memory     | Die pro Executor-Prozess zu verwendende Größe des Speichers (beispielsweise 1g, 2g).                                                                                        | Die Einstellung wird auf der Grundlage der Core- und der Aufgaben-Instance-Typen im Cluster konfiguriert.                                                                                                                                                                                                                                                        |
| spark.executor.cores      | Die für jeden Executor zu verwendende Anzahl der Kerne.                                                                                                                     | Die Einstellung wird auf der Grundlage der Core- und der Aufgaben-Instance-Typen im Cluster konfiguriert.                                                                                                                                                                                                                                                        |

| Einstellung                           | Beschreibung             | Wert                                                                                                                                                                                                                             |
|---------------------------------------|--------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <code>spark.executor.instances</code> | Die Anzahl der Executor. | Die Einstellung wird auf der Grundlage der Core- und der Aufgaben-Instance-Typen im Cluster konfiguriert. Festgelegt, wenn <code>spark.dynamicAllocation.enabled</code> nicht gleichzeitig explizit auf "true" festgelegt wurde. |

## Verhalten der Außerbetriebnahme von Knoten konfigurieren

Bei der Verwendung von Amazon-EMR-Version 5.9.0 oder höher, enthält Spark für Amazon EMR eine Reihe von Features, um sicherzustellen, dass Spark die Knotenbeendigung infolge einer manuellen Größenänderung oder einer Anforderung einer Auto-Scaling-Richtlinie zuverlässig ausführt. Amazon EMR implementiert einen Sperr-Mechanismus in Spark, der dem Außerbetriebnahme-Mechanismus von YARN übergeordnet ist. Dieser Mechanismus hilft sicherzustellen, dass keine neuen Aufgaben in einem Knoten geplant werden, der außer Betrieb genommen wird. Gleichzeitig werden bereits begonnene Aufgaben abgeschlossen. Darüber hinaus gibt es Funktionen, mit deren Hilfe Spark-Aufgaben schneller wiederhergestellt werden, wenn Shuffle-Blöcke bei der Beendigung eines Knotens verloren gehen. Der Neuberechnungsprozess wird früher ausgelöst und ist optimiert, um eine schnellere Neuberechnung mit weniger Stufenwiederholungen zu ermöglichen. Außerdem kann vermieden werden, dass Aufträge infolge von fehlenden Shuffle-Blöcken fehlschlagen.

### Important

Die Einstellung `spark.decommissioning.timeout.threshold` wurde in Amazon-EMR-Version 5.11.0 hinzugefügt, um die Ausfallsicherheit von Spark zu verbessern, wenn Sie Spot Instances verwenden. Wenn in früheren Versionen ein Knoten eine Spot Instance verwendet und die Instance aufgrund des Bid-Preises beendet wird, kann Spark die Beendigung möglicherweise nicht zuverlässig abschliessen. Aufträge können fehlschlagen und Shuffle-Neuberechnungen können eine geraume Zeit in Anspruch nehmen. Aus diesem Grund empfehlen wir die Nutzung von Version 5.11.0 oder höher, wenn Sie Spot Instances verwenden.

## Einstellungen für die Außerbetriebnahme von Spark Node

| Einstellung                                          | Beschreibung                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           | Standardwert      |
|------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------|
| <code>spark.blacklist.decommissioning.enabled</code> | Wenn die Einstellung auf <code>true</code> gesetzt ist, sperrt Spark die Knoten, die sich in YARN im <code>decommissioning</code> - Zustand befinden. Spark plant keine neuen Aufgaben auf Executors aus, die auf dem betreffenden Knoten ausgeführt werden. Bereits begonnene Aufgaben werden abgeschlossen.                                                                                                                                                                                                                                                                          | <code>true</code> |
| <code>spark.blacklist.decommissioning.timeout</code> | Die Zeitspanne, wie lange ein Knoten im Status <code>decommissioning</code> gesperrt wird. Standardmäßig ist dieser Wert auf eine Stunde festgelegt. Dies ist auch die Standardinstellung für <code>yarn.resourcemanager.decommissioning.timeout</code> . Um sicherzustellen, dass ein Knoten über den gesamten Außerbetriebnahme-Zeitraum gesperrt wird, setzen Sie diesen Wert größer oder gleich <code>yarn.resourcemanager.decommissioning.timeout</code> . Nachdem die Frist für die Außerbetriebnahme abgelaufen ist, wird der Knoten in den Zustand <code>decommissioned</code> | <code>1h</code>   |

| Einstellung | Beschreibung                                                                                                                                                                                                                                                                  | Standardwert |
|-------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------|
|             | überführt, und Amazon EMR kann die EC2-Instance des Knotens beenden. Wenn noch Aufgaben ausgeführt werden, nachdem die Zeitbeschränkung abgelaufen ist, gehen diese verloren oder werden abgebrochen und auf Executors, die auf anderen Knoten ausgeführt werden, verschoben. |              |



| Einstellung                                          | Beschreibung                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 | Standardwert |
|------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------|
| <code>spark.decommissioning.timeout.threshold</code> | Verfügbar in Amazon-EMR-Version 5.11.0 oder höher.<br>In Sekunden angegeben.<br>. Wenn ein Knoten in den Status „Außerbetriebnahme“ übergeht und der Host ihn innerhalb der durch diesen Wert festgelegten Frist außer Betrieb nimmt, sperrt Amazon EMR den Knoten nicht nur, sondern bereinigt auch den Host-Status (wie in <code>spark.resourceManager.cleanupExpiredHost</code> angegeben), ohne darauf zu warten, dass der Knoten in den Status „Außerbetriebnahme“ übergeht. Auf diese Weise kann Spark das Beenden von Spot-Instances besser verarbeiten, weil Spot-Instances innerhalb eines Timeouts von 20 Sekunden außer Betrieb genommen werden, unabhängig von dem Wert von <code>yarn.resourcemanager.decommissioning.timeout</code> , sodass andere Knoten möglicherweise nicht genügend Zeit haben, Shuffle-Dateien zu lesen. | 20s          |

| Einstellung                                                       | Beschreibung                                                                                                                                                                                                                                                                                                                                                                                        | Standardwert      |
|-------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------|
| <code>spark.resourceManager.cleanupExpiredHost</code>             | Wenn die Einstellung auf <code>true</code> gesetzt ist, löscht Spark alle zwischengespeicherten Daten und Shuffle-Blöcke, die in Executors in Knoten gespeichert sind, die sich im Zustand <code>decommissioned</code> befinden. Dadurch wird der Wiederherstellungsprozess beschleunigt.                                                                                                           | <code>true</code> |
| <code>spark.stage.attempt.ignoreOnDecommissionFetchFailure</code> | Wenn die Einstellung auf <code>true</code> gesetzt ist, wird verhindert, dass in Spark Stufen und letztendlich Aufträge wegen zu vieler Abrufe von außer Betrieb genommenen Knoten fehlschlagen. Fehler beim Abrufen von Shuffle-Blöcken von einem Knoten im Zustand <code>decommissioned</code> werden nicht zu der maximal zulässigen Anzahl von aufeinander folgenden Abruffehlern hinzugezählt. | <code>true</code> |

## ThriftServer Spark-Umgebungsvariable

Spark legt die Hive-Thrift-Server-Port-Umgebungsvariable `HIVE_SERVER2_THRIFT_PORT` auf "10001" fest.

## Ändern von Spark-StandardEinstellungen

Sie ändern die Standardwerte in `spark-defaults.conf` mit der Konfigurationsklassifizierung `spark-defaults` oder der Einstellung `maximizeResourceAllocation` in der Konfigurationsklassifizierung `spark`.

In den folgenden Verfahren wird gezeigt, wie Sie Einstellungen mithilfe der CLI oder der Konsole ändern.

Erstellen Sie einen Cluster mithilfe der CLI mit der Einstellung „`spark.executor.memory`“ auf „2G“ wie folgt

- Erstellen Sie einen Cluster mit installiertem Spark und `spark.executor.memory` auf 2G gesetzt, indem Sie den folgenden Befehl eingeben, der auf die Datei `myConfig.json` verweist, die in Amazon S3 gespeichert ist.

```
aws emr create-cluster --release-label emr-7.1.0 --applications Name=Spark \  
--instance-type m5.xlarge --instance-count 2 --service-role EMR_DefaultRole_V2  
--ec2-attributes InstanceProfile=EMR_EC2_DefaultRole --configurations https://  
s3.amazonaws.com/mybucket/myfolder/myConfig.json
```

### Note

Linux-Zeilenfortsetzungszeichen (`\`) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (`^`).

`myConfig.json`:

```
[  
  {  
    "Classification": "spark-defaults",  
    "Properties": {  
      "spark.executor.memory": "2G"  
    }  
  }  
]
```

Erstellen Sie einen Clustermithilfe der Konsole mit der Einstellung „spark.executor.memory“ auf „2G“ wie folgt

1. Navigieren Sie zur neuen Amazon-EMR-Konsole und wählen Sie in der Seitennavigation die Option Zur alten Konsole wechseln aus. Weitere Informationen darüber, was Sie erwartet, wenn Sie zur alten Konsole wechseln, finden Sie unter [Verwenden der alten Konsole](#).
2. Wählen Sie Create Cluster (Cluster erstellen) und Go to advanced options (Zu erweiterten Optionen) aus.
3. Wählen Sie Spark aus.
4. Lassen Sie unter Edit software settings (Softwareeinstellungen bearbeiten) die Option Enter configuration (Konfiguration eingeben) ausgewählt und nehmen Sie folgende Konfiguration vor:

```
classification=spark-defaults,properties=[spark.executor.memory=2G]
```

5. Wählen Sie andere Optionen aus, dann und anschließend Create cluster (Cluster erstellen).

Zum Einstellen maximizeResourceAllocation

- Erstellen Sie einen Cluster, auf dem Spark installiert und auf true maximizeResourceAllocation gesetzt ist AWS CLI, indem Sie die in Amazon S3 gespeicherte Datei referenzierend verwenden. myConfig.json

```
aws emr create-cluster --release-label emr-7.1.0 --applications Name=Spark \  
--instance-type m5.xlarge --instance-count 2 --service-role EMR_DefaultRole_V2 \  
--ec2-attributes InstanceProfile=EMR_EC2_DefaultRole --configurations https://  
s3.amazonaws.com/mybucket/myfolder/myConfig.json
```

#### Note

Linux-Zeilenumbruchzeichen (\) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (^).

myConfig.json:

```
[  
{
```

```
"Classification": "spark",
"Properties": {
  "maximizeResourceAllocation": "true"
}
]
```

### Note

Ab Amazon-EMR-Version 5.21.0 können Sie Cluster-Konfigurationen überschreiben und zusätzliche Konfigurationsklassifikationen für jede Instance-Gruppe in einem ausgeführten Cluster angeben. Dazu verwenden Sie die Amazon EMR-Konsole, das AWS Command Line Interface (AWS CLI) oder das AWS SDK. Weitere Informationen finden Sie unter [Angeben einer Konfiguration für eine Instance-Gruppe in einem aktiven Cluster](#).

## Migration von Apache Log4j 1.x zu Log4j 2.x

[Apache-Spark-Versionen](#) 3.2.x und frühere Versionen verwenden den älteren Apache Log4j 1.x und die `log4j.properties`-Datei, um Log4j in Spark-Prozessen zu konfigurieren. Apache-Spark-Versionen 3.3.0 und höher verwenden Apache Log4j 2.x und die Datei `log4j2.properties`, um Log4j in Spark-Prozessen zu konfigurieren.

Wenn Sie Apache Spark Log4j mit einer Amazon-EMR-Version vor 6.8.0 konfiguriert haben, müssen Sie die alte `spark-log4j`-Konfigurationsklassifizierung entfernen und zur `spark-log4j2`-Konfigurationsklassifizierung und zum Schlüsselformat migrieren, bevor Sie auf Amazon EMR 6.8.0 oder höher aktualisieren können. Die `spark-log4j-Legacy`-Klassifizierung führt dazu, dass die Clustererstellung in Amazon-EMR-Versionen 6.8.0 und höher mit einem `ValidationException`-Fehler fehlschlägt. Ein Fehler im Zusammenhang mit der Log4j-Inkompatibilität wird Ihnen nicht in Rechnung gestellt. Sie müssen jedoch die Klassifizierung für die veraltete `spark-log4j`-Konfiguration entfernen, um fortzufahren.

Weitere Informationen zur Migration von Apache Log4j 1.x zu Log4j 2.x finden Sie im [Apache Log4j Migration Benutzerhandbuch](#) und im [Spark Log4j 2 Template](#) auf Github.

**Note**

Bei Amazon EMR verwendet Apache Spark eine `log4j2.properties`-Datei anstelle der im [Apache-Log4j-Migration-Benutzerhandbuch](#) beschriebenen `.xml`-Datei. Außerdem empfehlen wir nicht, die `Log4j-1.x-Bridge`-Methode für die Konvertierung nach `Log4j 2.x` zu verwenden.

## Die Leistung von Spark optimieren

Amazon EMR bietet mehrere Features für die Leistungsoptimierung von Spark. In diesem Thema werden die einzelnen Optimierungsfunktionen im Detail erläutert.

Weitere Informationen zum Festlegen der Spark-Konfiguration finden Sie unter [Konfigurieren von Spark](#).

### Adaptive Abfrageausführung

Die adaptive Abfrageausführung ist ein Framework zur Neuoptimierung von Abfrageplänen auf der Grundlage von Laufzeitstatistiken. Ab Amazon EMR 5.30.0 sind die folgenden Optimierungen für die adaptive Abfrageausführung von Apache Spark 3 auf Apache-Amazon-EMR-Laufzeit für Spark 2 verfügbar.

- Adaptive Join-Konvertierung
- Adaptives Zusammenführen von Shuffle-Partitionen

#### Adaptive Join-Konvertierung

Die adaptive Join-Konvertierung verbessert die Abfrageleistung, indem `sort-merge-join` Operationen in Operationen umgewandelt werden, `broadcast-hash-joins` die auf der Laufzeitgröße der Abfragephasen basieren. `Broadcast-hash-joins` schneidet tendenziell besser ab, wenn eine Seite des Joins klein genug ist, um seine Ausgabe effizient an alle Executoren zu verteilen. Dadurch entfällt die Notwendigkeit, den Austausch zu mischen und beide Seiten des Joins zu sortieren. Die adaptive Join-Konvertierung erweitert das Spektrum der Fälle, in denen Spark automatisch ausgeführt wird.

`broadcast-hash-joins`

Dieses Feature ist standardmäßig aktiviert. Sie kann deaktiviert werden, indem `spark.sql.adaptive.enabled` auf `false` gesetzt wird, wodurch auch der

adaptive Abfrageausführungsrahmen deaktiviert wird. Spark entscheidet, a in a sort-merge-join zu konvertieren, broadcast-hash-join wenn die Laufzeitgrößenstatistik einer der Join-Seiten den Standardwert von 10.485.760 Byte (10 MiB) nicht überschreitet `spark.sql.autoBroadcastJoinThreshold`.

## Adaptives Zusammenführen von Shuffle-Partitionen

Das adaptive Zusammenführen von Shuffle-Partitionen verbessert die Abfrageleistung, indem kleine zusammenhängende Shuffle-Partitionen zusammengeführt werden, um den Mehraufwand zu vermeiden, der durch zu viele kleine Aufgaben entsteht. Auf diese Weise können Sie eine höhere Anzahl von anfänglichen Shuffle-Partitionen konfigurieren, die dann zur Laufzeit auf eine bestimmte Größe reduziert werden, wodurch sich die Wahrscheinlichkeit erhöht, dass Shuffle-Partitionen gleichmäßiger verteilt werden.

Dieses Feature ist standardmäßig aktiviert, sofern nicht ausdrücklich `spark.sql.shuffle.partitions` festgelegt ist. Sie kann aktiviert werden, indem Sie `spark.sql.adaptive.coalescePartitions.enabled` auf `true` einstellen. Sowohl die anfängliche Anzahl der Shuffle-Partitionen als auch die Größe der Zielpartition können mit den Eigenschaften `spark.sql.adaptive.coalescePartitions.minPartitionNum` und `spark.sql.adaptive.advisoryPartitionSizeInBytes` eingestellt werden. In der folgenden Tabelle finden Sie weitere Informationen zu den zugehörigen Spark-Eigenschaften für dieses Feature.

## Adaptive Koaleszenzpartitioneigenschaften von Spark

| Eigenschaft                                                | Standardwert                                                                           | Beschreibung                                                                                                                                                                                                                                                                                |
|------------------------------------------------------------|----------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <code>spark.sql.adaptive.coalescePartitions.enabled</code> | „Wahr“, sofern <code>spark.sql.shuffle.partitions</code> nicht explizit festgelegt ist | Wenn „Wahr“ und <code>spark.sql.adaptive.enabled</code> den Wert „Wahr“ hat, fügt Spark zusammenhängende Shuffle-Partitionen entsprechend der Zielgröße (angegeben durch <code>spark.sql.adaptive.advisoryPartitionSizeInBytes</code> ) zusammen, um zu viele kleine Aufgaben zu vermeiden. |

| Eigenschaft                                                            | Standardwert | Beschreibung                                                                                                                                                                                                                                                          |
|------------------------------------------------------------------------|--------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <code>spark.sql.adaptive.advisoryPartitionSizeInBytes</code>           | 64 MB        | Die empfohlene Größe der Shuffle-Partition beim Zusammenführen. Diese Konfiguration wirkt sich nur aus, wenn <code>spark.sql.adaptive.enabled</code> und <code>spark.sql.adaptive.coalescePartitions.enabled</code> beide gleichzeitig <code>true</code> sind.        |
| <code>spark.sql.adaptive.coalescePartitions.minPartitionNum</code>     | 25           | Die Mindestanzahl der Shuffle-Partitionen nach dem Zusammenführen. Diese Konfiguration wirkt sich nur aus, wenn <code>spark.sql.adaptive.enabled</code> und <code>spark.sql.adaptive.coalescePartitions.enabled</code> beide gleichzeitig <code>true</code> sind.     |
| <code>spark.sql.adaptive.coalescePartitions.initialPartitionNum</code> | 1000         | Die anfängliche Anzahl von Shuffle-Partitionen vor dem Zusammenführen. Diese Konfiguration wirkt sich nur aus, wenn <code>spark.sql.adaptive.enabled</code> und <code>spark.sql.adaptive.coalescePartitions.enabled</code> beide gleichzeitig <code>true</code> sind. |



## Dynamische Partitionsbereinigung

Die dynamische Partitionsbereinigung verbessert die Auftragsleistung durch eine sorgfältigere Auswahl der spezifischen Partitionen in einer Tabelle, die für eine bestimmte Abfrage gelesen und verarbeitet werden müssen. Indem Sie die Datenmenge reduzieren, die für die Ausführung eines Auftrags gelesen und verarbeitet werden muss, können Sie viel Zeit sparen. In Amazon EMR 5.26.0 ist dieses Feature standardmäßig aktiviert. In Amazon EMR 5.24.0 und 5.25.0 können Sie dieses Feature aktivieren, indem Sie die Spark-Eigenschaft `spark.sql.dynamicPartitionPruning.enabled` innerhalb von Spark oder beim Erstellen von Clustern festlegen.

### Dynamisches Löschen von Partitionen in Spark, Partitionseigenschaften

| Eigenschaft                                                                    | Standardwert      | Beschreibung                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
|--------------------------------------------------------------------------------|-------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <code>spark.sql.dynamicPartitionPruning.enabled</code>                         | <code>true</code> | Wenn der Wert wahr ist, aktivieren Sie das dynamische Bereinigen von Partitionen.                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
| <code>spark.sql.optimizer.dynamicPartitionPruning.enforceBroadcastReuse</code> | <code>true</code> | Wenn <code>true</code> , führt Spark vor der Ausführung der Abfrage eine Defensivprüfung durch, um sicherzustellen, dass die Wiederverwendung von Broadcast-Austauschen in dynamischen Bereinigungsfiltern nicht durch spätere Vorbereitungsregeln, wie z. B. benutzerdefinierte Spaltenregeln, beeinträchtigt wird. Wenn die Wiederverwendung nicht funktioniert und diese Konfiguration <code>true</code> ist, entfernt Spark die betroffenen dynamischen Bereinigungsfilter, um Leistungs- und Korrektheitsprobleme zu vermeiden. |

| Eigenschaft | Standardwert | Beschreibung                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
|-------------|--------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|             |              | Korrektheitsprobleme können auftreten, wenn der Broadcast-Austausch des dynamischen Bereinigungsfilters zu unterschiedlichen, inkonsistenten Ergebnissen führt, als der Broadcast-Austausch des entsprechenden Join-Vorgangs. Das Einstellen dieser Konfiguration auf <code>false</code> sollte mit Vorsicht erfolgen. Dadurch können Szenarien umgangen werden, z. B. wenn die Wiederverwendung durch benutzerdefinierte Spaltenregeln gestört wird. Wenn Adaptive Abfrageausführung aktiviert ist, wird die Wiederverwendung von Broadcasts immer erzwungen. |

Diese Optimierung verbessert die vorhandenen Funktionen von Spark 2.4.2, das nur die Weitergabe statischer Prädikate unterstützt, um diese zu geplanten Zeiten aufzulösen.

Im Folgenden finden Sie Beispiele für die Weitergabe statischer Prädikate in Spark 2.4.2.

```
partition_col = 5  
  
partition_col IN (1,3,5)  
  
partition_col between 1 and 3  
  
partition_col = 1 + 3
```

Die dynamische Partitionsbereinigung ermöglicht es der Spark-Engine, während der Laufzeit dynamisch abzuleiten, welche Partitionen gelesen werden müssen und welche problemlos

eliminiert werden können. Die folgende Abfrage beinhaltet beispielsweise zwei Tabellen: die Tabelle `store_sales`, die den Gesamtumsatz aller Geschäfte enthält und nach Regionen partitioniert ist, sowie die Tabelle `store_regions`, die für die einzelnen Länder eine Zuweisung nach Regionen enthält. Die Tabellen enthalten Daten für Geschäfte, die auf der ganzen Welt verteilt sind, allerdings benötigen wir nur die Daten für Nordamerika.

```
select ss.quarter, ss.region, ss.store, ss.total_sales
from store_sales ss, store_regions sr
where ss.region = sr.region and sr.country = 'North America'
```

Ohne die dynamische Partitionsbereinigung, liest diese Anfrage alle Regionen, bevor Sie die Untergruppe an Regionen herausfiltert, die mit den Ergebnissen der Unterabfrage übereinstimmen. Mit der dynamischen Partitionsbereinigung liest und verarbeitet diese Abfrage nur die Partitionen für die Regionen, die in der Unterabfrage zurückgegeben wurden. Da weniger Daten im Speicher gelesen und weniger Datensätze verarbeitet werden müssen, spart dies Zeit und Ressourcen.

## Abflachen skalarer Unterabfragen

Diese Optimierung verbessert die Leistung von Abfragen, die in der gleichen Tabelle skalare Unterabfragen ausführen. In Amazon EMR 5.26.0 ist dieses Feature standardmäßig aktiviert. In Amazon EMR 5.24.0 und 5.25.0 können Sie es aktivieren, indem Sie die Spark-Eigenschaft `spark.sql.optimizer.flattenScalarSubqueriesWithAggregates.enabled` innerhalb von Spark oder beim Erstellen von Clustern festlegen. Wenn diese Eigenschaft auf „true“ festgelegt ist, flacht der Abfrageoptimierer aggregierte skalare Unterabfragen, die, wenn möglich, dieselbe Relation verwenden, ab. Die skalaren Unterabfragen werden abgeflacht, indem alle in der Unterabfrage vorhandenen Prädikate an die Aggregationsfunktionen weitergegeben werden. Im Anschluss daran wird pro Relation eine Aggregation mit allen Aggregatfunktionen ausgeführt.

Im Folgenden finden Sie ein Beispiel für eine Abfrage, die von dieser Optimierung profitiert.

```
select (select avg(age) from students /* Subquery 1 */
       where age between 5 and 10) as group1,
       (select avg(age) from students /* Subquery 2 */
       where age between 10 and 15) as group2,
       (select avg(age) from students /* Subquery 3 */
       where age between 15 and 20) as group3
```

Die Optimierung schreibt die vorherige Abfrage folgendermaßen um:

```
select c1 as group1, c2 as group2, c3 as group3
from (select avg (if(age between 5 and 10, age, null)) as c1,
          avg (if(age between 10 and 15, age, null)) as c2,
          avg (if(age between 15 and 20, age, null)) as c3 from students);
```

Beachten Sie, dass die umgeschriebene Abfrage die Studententabelle nur einmal liest und die Prädikate der drei Unterabfragen per Push-Verfahren in die avg-Funktion weitergegeben werden.

## DISTINCT vor INTERSECT

Diese Optimierung verbessert Joins, wenn INTERSECT verwendet wird. In Amazon EMR 5.26.0 ist dieses Feature standardmäßig aktiviert. In Amazon EMR 5.24.0 und 5.25.0 können Sie es aktivieren, indem Sie die Spark-Eigenschaft `spark.sql.optimizer.distinctBeforeIntersect.enabled` innerhalb von Spark oder beim Erstellen von Clustern festlegen. Abfragen mit INTERSECT werden automatisch so konvertiert, dass sie einen linken halben Join verwenden. Wenn diese Eigenschaft auf true gesetzt ist, überträgt der Abfrageoptimierer den DISTINCT-Operator an die untergeordneten Objekte von INTERSECT, wenn er feststellt, dass der DISTINCT-Operator die linke Semi-Verknüpfung zu einem statt zu a machen kann. BroadcastHashJoin SortMergeJoin

Im Folgenden finden Sie ein Beispiel für eine Abfrage, die von dieser Optimierung profitiert.

```
(select item.brand brand from store_sales, item
  where store_sales.item_id = item.item_id)
intersect
(select item.brand cs_brand from catalog_sales, item
  where catalog_sales.item_id = item.item_id)
```

Wenn die Eigenschaft `spark.sql.optimizer.distinctBeforeIntersect.enabled`, nicht aktiviert ist, wird die Abfrage folgendermaßen neu geschrieben.

```
select distinct brand from
  (select item.brand brand from store_sales, item
    where store_sales.item_id = item.item_id)
left semi join
  (select item.brand cs_brand from catalog_sales, item
    where catalog_sales.item_id = item.item_id)
on brand <=> cs_brand
```

Wenn die Eigenschaft `spark.sql.optimizer.distinctBeforeIntersect.enabled`, nicht aktiviert ist, wird die Abfrage folgendermaßen neu geschrieben.

```
select brand from
  (select distinct item.brand brand from store_sales, item
    where store_sales.item_id = item.item_id)
left semi join
  (select distinct item.brand cs_brand from catalog_sales, item
    where catalog_sales.item_id = item.item_id)
on brand <=> cs_brand
```

## Bloomfilter für Joins

Durch diese Optimierung kann die Leistung einiger Joins verbessert werden, da eine Seite eines Joins mit einem [Bloomfilter](#) vorgefiltert wird, der aus den Werten von der anderen Seite des Joins generiert wird. In Amazon EMR 5.26.0 ist dieses Feature standardmäßig aktiviert. In Amazon EMR 5.25.0 können Sie dieses Feature aktivieren, indem Sie die Spark-Eigenschaft `spark.sql.bloomFilterJoin.enabled` innerhalb von Spark oder beim Erstellen von Clustern auf `true` setzen.

Im Folgenden finden Sie eine Beispielabfrage, für die die Anwendung eines Bloomfilters geeignet wäre.

```
select count(*)
from sales, item
where sales.item_id = item.id
and item.category in (1, 10, 16)
```

Wenn diese Funktion aktiviert ist, wird der Bloomfilter aus allen Artikel-IDs erstellt, deren Kategorie sich innerhalb des Kategoriensatzes befindet, der abgefragt wird. Beim Scannen der Tabelle SALES wird der Bloomfilter verwendet, um zu ermitteln, welche Verkäufe für Artikel gelten, die sich definitiv nicht im vom Bloomfilter definierten Satz befinden. So können diese ermittelten Verkäufe so früh wie möglich herausgefiltert werden.

## Optimierte Join-Neuanordnung

Durch diese Optimierung kann die Abfrageleistung verbessert werden, indem bei Tabellen mit Filtern die Joins neu angeordnet werden. In Amazon EMR 5.26.0 ist dieses Feature standardmäßig aktiviert. In Amazon EMR 5.25.0, können Sie dieses Feature aktivieren, indem Sie den Spark-Konfigurationsparameter `spark.sql.optimizer.sizeBasedJoinReorder.enabled` auf „wahr“

setzen. Das Standardverhalten von Spark ist, dass Tabellen von links nach rechts verknüpft werden, wie in der Abfrage aufgeführt. Dabei kann die Chance verpasst werden, zuerst kleinere Joins mit Filtern auszuführen, was später teureren Joins zugute kommen würde.

Bei der Beispielabfrage unten werden alle zurückgegebenen Artikel aus allen Läden in einem Land ausgewertet. Ohne optimierte Join-Neuanordnung verknüpft Spark zuerst die zwei großen Tabellen `store_sales` und `store_returns` und verknüpft sie dann mit `store` und schließlich mit `item`.

```
select ss.item_value, sr.return_date, s.name, i.desc,
from store_sales ss, store_returns sr, store s, item i
where ss.id = sr.id and ss.store_id = s.id and ss.item_id = i.id
and s.country = 'USA'
```

Bei der optimierten Join-Neuanordnung verknüpft Spark zuerst `store_sales` mit `store`, da `store` über einen Filter verfügt und kleiner ist als `store_returns` und `broadcastable`. Danach führt Spark einen Join mit `store_returns` und schließlich mit `item` durch. Wenn `item` über einen Filter verfügen würde und sendungsfähig wäre, würde es sich ebenfalls für die Neuanordnung eignen, was dazu führen würde, dass `store_sales` mit `store`, dann mit `item` und schließlich mit `store_returns` verknüpft würde.

## Caching von Ergebnisfragmenten in Spark

Amazon EMR 6.6.0 und höher enthalten das optionale Feature zum Caching von Ergebnisfragmenten in Spark, mit dem Ergebnisfragmente automatisch zwischengespeichert werden. Diese Ergebnisfragmente sind Teile von Ergebnissen aus Teilbäumen von Abfragen, die in einem Amazon-S3-Bucket Ihrer Wahl gespeichert sind. Die gespeicherten Fragmente der Abfrageergebnisse werden bei nachfolgenden Abfrageausführungen wiederverwendet, was zu schnelleren Abfragen führt.

Beim Caching von Ergebnisfragmenten werden Ihre Spark-SQL-Abfragen analysiert und geeignete Ergebnisfragmente an Ihrem angegebenen S3-Speicherort zwischengespeichert. Bei nachfolgenden Abfrageläufen werden die verwendbaren Fragmente der Abfrageergebnisse automatisch erkannt und von S3 abgerufen. Das Caching von Ergebnisfragmenten unterscheidet sich vom Zwischenspeichern von Ergebnismengen, bei dem nachfolgende Abfragen exakt mit der ursprünglichen Abfrage übereinstimmen müssen, um Ergebnisse aus dem Cache zurückzugeben. Wenn es für Abfragen verwendet wird, die wiederholt auf eine statische Teilmenge Ihrer Daten abzielen, beschleunigt das Caching von Ergebnisfragmenten die Leistung erheblich.

Stellen Sie sich die folgende Abfrage vor, die Bestellungen bis zum Jahr 2022 zählt:

```
select
  l_returnflag,
  l_linestatus,
  count(*) as count_order
from
  lineitem
where
  l_shipdate <= current_date
  and year(l_shipdate) == '2022'
group by
  l_returnflag,
  l_linestatus
```

Im Laufe der Zeit muss diese Abfrage täglich ausgeführt werden, um den Gesamtumsatz des Jahres zu melden. Ohne das Caching von Ergebnisfragmenten müssen die Ergebnisse für alle Tage des Jahres täglich neu berechnet werden. Die Abfrage wird mit der Zeit langsamer und am Jahresende, wenn die Ergebnisse aller 365 Tage neu berechnet werden müssen, am langsamsten.

Wenn Sie das Caching von Ergebnisfragmenten aktivieren, verwenden Sie Ergebnisse für alle vorherigen Tage des Jahres aus dem Cache. An jedem Tag muss das Feature nur die Ergebnisse eines Tages neu berechnen. Nachdem das Feature das Ergebnisfragment berechnet hat, speichert das Feature das Fragment im Cache. Dadurch sind die Cache-aktivierten Abfragezeiten schnell und sie bleiben bei jeder nachfolgenden Abfrage konstant.

## Caching von Ergebnisfragmenten in Spark aktivieren

Zum Aktivieren von Caching von Ergebnisfragmenten in Spark führen Sie die folgenden Schritte durch:

1. Erstellen Sie einen Cache-Bucket in Amazon S3 und autorisieren Sie den Lese-/Schreibzugriff für EMRFS. Weitere Informationen finden Sie unter [Zugriff auf EMRFS-Daten in Amazon S3 genehmigen](#).
2. Stellen Sie die Amazon-EMR-Spark-Konfiguration ein, um das Feature zu aktivieren.

```
spark.subResultCache.enabled = true
spark.subResultCache.fs.root.path = s3://DOC-EXAMPLE-BUCKET/cache_dir/
```

3. Aktivieren Sie das S3-Lebenszyklusmanagement für den Bucket, um Cache-Dateien automatisch zu bereinigen.

4. Optional können Sie die `maxBufferSize` Eigenschaften `reductionRatioThreshold` und konfigurieren, um die Funktion weiter zu optimieren.

```
spark.sql.subResultCache.reductionRatioThreshold  
spark.sql.subResultCache.maxBufferSize
```

## Einige Hinweise zur Verwendung von Caching von Ergebnisfragmenten

Die Kosteneinsparungen, wenn Sie Ergebnisse verwenden, die bereits in Amazon S3 zwischengespeichert wurden, anstatt sie neu zu berechnen, steigen mit der Häufigkeit, mit der dieselben zwischengespeicherten Ergebnisse verwendet werden können. Abfragen mit umfangreichen Tabellenscans, gefolgt von Filtern oder Hash-Aggregationen, die die Ergebnisgröße um den Faktor 8 reduzieren (d. h. ein Verhältnis von mindestens 8:1 bei Eingabegröße zu Ergebnissen), profitieren am meisten von diesem Feature. Je größer das Reduktionsverhältnis zwischen Eingabe und Ergebnissen ist, desto größer ist der Kostenvorteil. Abfragen mit kleineren Reduktionsraten, die jedoch teure Rechenschritte zwischen dem Tabellenscan und Filtern oder Aggregationen enthalten, werden ebenfalls davon profitieren, sofern die Kosten für die Erstellung der Ergebnisse höher sind als die Kosten für das Abrufen der Ergebnisse aus Amazon S3. Standardmäßig wird das Caching von Ergebnisfragmenten nur wirksam, wenn erkannt wird, dass das Reduktionsverhältnis mindestens 8:1 beträgt.

Wenn Ihre Abfragen zwischengespeicherte Ergebnisse wiederholt wiederverwenden, sind die Vorteile dieses Features am größten. Rollende und inkrementelle Fensterabfragen sind gute Beispiele. Beispielsweise müsste bei einer 30-tägigen rollenden Fenster-Abfrage, die bereits 29 Tage lang ausgeführt wurde, nur 1/30 der Zieldaten aus der ursprünglichen Eingabequelle abgerufen werden, und es würden zwischengespeicherte Ergebnisfragmente für die 29 vorherigen Tage verwendet. Eine inkrementelle Fensterabfrage wäre sogar noch besser, da der Beginn des Fensters unverändert bleibt: Bei jedem Aufruf der Abfrage muss ein kleinerer Prozentsatz der Verarbeitung aus der Eingabequelle gelesen werden.

Im Folgenden finden Sie weitere Überlegungen zur Verwendung von Caching von Ergebnisfragmenten:

- Abfragen, die nicht auf dieselben Daten mit denselben Abfragefragmenten abzielen, haben eine niedrige Cache-Trefferrate und profitieren daher nicht von diesem Feature.



- Abfragen mit niedrigen Reduktionsraten, die keine teuren Rechenschritte enthalten, führen zu zwischengespeicherten Ergebnissen, deren Lesen ungefähr so teuer ist wie die ursprüngliche Verarbeitung.
- Die erste Abfrage weist aufgrund der Kosten für das Schreiben in den Cache immer eine geringfügige Regression auf.
- Das Feature zum Caching von Ergebnisfragmenten funktioniert ausschließlich mit Parquet-Dateien. Andere Dateiformate werden nicht unterstützt.
- Das Feature Caching von Ergebnisfragmenten für das Zwischenspeichern von Ergebnisfragmenten versucht nur, Scans mit Datei-Split-Größen von 128 MB oder mehr zwischenzuspeichern. Bei der Standardkonfiguration von Spark ist das Caching von Ergebnisfragmenten deaktiviert, wenn die Scangröße (Gesamtgröße aller gescannten Dateien) geteilt durch die Anzahl der Executor-Cores weniger als 128 MB beträgt. Wenn eine der unten aufgeführten Spark-Konfigurationen festgelegt ist, beträgt die Größe der aufgeteilten Datei:

```
min(maxPartitionBytes, max(openCostInBytes, scan size / minPartitionNum))
```

- `spark.sql.leafNodeDefaultParallelismus` (Standardwert ist `spark.default.parallelism`)
- `spark.sql.files.minPartitionNum` (Der Standardwert ist `spark.sql.leafNodeDefaultParallelität`)
- `spark.sql.files.openCostInByte`
- `spark.sql.dateien.maxPartitionBytes`
- Das Feature zum Caching von Ergebnisfragmenten speichert die Granularität der RDD-Partition im Cache. Das zuvor beschriebene Reduktionsverhältnis, das standardmäßig 8:1 beträgt, wird pro RDD-Partition bewertet. Bei Workloads mit Reduktionsraten pro RDD, die sowohl größer als auch kleiner als 8:1 sind, kann es zu geringeren Leistungsvorteilen kommen als bei Workloads mit Reduktionsraten pro RDD, die durchweg unter 8:1 liegen.
- Das Feature zum Caching von Ergebnisfragmenten verwendet standardmäßig einen Schreibpuffer von 16 MB für jede RDD-Partition, die zwischengespeichert wird. Wenn mehr als 16 MB pro RDD-Partition zwischengespeichert werden, können die Kosten für die Feststellung, dass kein Schreibvorgang möglich ist, zu Leistungsregressionen führen.
- Standardmäßig versucht Caching von Ergebnisfragmenten zwar nicht, Ergebnisse einer RDD-Partition mit einem Reduktionsverhältnis von weniger als 8:1 zwischenzuspeichern und seinen Schreibpuffer auf 16 MB begrenzt, aber beide Werte können über die folgenden Konfigurationen eingestellt werden:

```
spark.sql.subResultCache.reductionRatioThreshold (default: 8.0)
```

```
spark.sql.subResultCache.maxBufferSize (default: 16MB, max: 64MB)
```

- Mehrere Cluster, die dieselbe Amazon-EMR-Version verwenden, können sich denselben Cache-Speicherort teilen. Um die Richtigkeit der Ergebnisse sicherzustellen, verwendet Caching von Ergebnisfragmenten keine Cache-Ergebnisse, die von verschiedenen Versionen von Amazon EMR geschrieben wurden.
- Das Zwischenspeichern von Ergebnisfragmenten wird automatisch für Spark Streaming-Anwendungsfälle deaktiviert oder wenn RecordServer Apache Ranger oder verwendet AWS Lake Formation wird.
- Beim Lesen/Schreiben des Ergebnisfragment-Cache werden EMRFS- und Amazon-S3-Buckets verwendet. CSE/ SSE S3/SSE KMS-Verschlüsselung wird unterstützt.

## Den Nvidia-RAPIDS-Accelerator für Apache Spark verwenden

Mit Amazon-EMR-Version 6.2.0 und höher können Sie das Plugin [RAPIDS Accelerator für Apache Spark](#) von Nvidia verwenden, um Spark mithilfe von EC2-GPU (Graphics Processing Unit)-Instance-Typen zu beschleunigen. RAPIDS Accelerator beschleunigt Ihre Apache-Spark-3.0-Pipelines für die Datenwissenschaft ohne Codeänderungen mit der GPU, beschleunigt die Datenverarbeitung und das Modelltraining und senkt gleichzeitig die Infrastrukturkosten erheblich.

Die folgenden Abschnitte führen Sie durch die Konfiguration Ihres EMR-Clusters für die Verwendung des Spark-Rapids-Plugins für Spark.

### Instance-Typen auswählen

Um das Nvidia-Spark-RAPIDS-Plugin für Spark verwenden zu können, müssen die Kern- und Aufgaben-Instance-Gruppen EC2-GPU-Instance-Typen verwenden, die die [Hardwareanforderungen von Spark-RAPIDS](#) erfüllen. Eine vollständige Liste der von Amazon EMR unterstützten GPU-Instance-Typen finden Sie unter [Unterstützte Instance-Typen](#) im Verwaltungshandbuch für Amazon EMR. Der Instance-Typ für die primäre Instance-Gruppe kann entweder GPU- oder Nicht-GPU-Typen sein, ARM-Instance-Typen werden jedoch nicht unterstützt.

### Anwendungskonfigurationen für Ihren Cluster einrichten

1. Amazon EMR aktivieren, um die Plugins auf Ihrem neuen Cluster zu installieren

Zum Installieren von Plugins geben Sie beim Erstellen Ihres Clusters die folgende Konfiguration an:

```
{
  "Classification":"spark",
  "Properties":{
    "enableSparkRapids":"true"
  }
}
```

## 2. YARN für die Verwendung von GPU konfigurieren

Einzelheiten zur Verwendung von GPU auf YARN finden Sie unter [Verwenden von GPU auf YARN](#) in der Apache-Hadoop-Dokumentation. Die folgenden Beispiele zeigen YARN-Beispielkonfigurationen für die Amazon-EMR-Versionen 6.x und 7.x:

### Amazon EMR 7.x

#### Beispiel für eine YARN-Konfiguration für Amazon EMR 7.x

```
{
  "Classification":"yarn-site",
  "Properties":{
    "yarn.nodemanager.resource-plugins":"yarn.io/gpu",
    "yarn.resource-types":"yarn.io/gpu",
    "yarn.nodemanager.resource-plugins.gpu.allowed-gpu-devices":"auto",
    "yarn.nodemanager.resource-plugins.gpu.path-to-discovery-executables":"/usr/
bin",
    "yarn.nodemanager.linux-container-executor.cgroups.mount":"true",
    "yarn.nodemanager.linux-container-executor.cgroups.mount-path":"/spark-
rapids-cgroup",
    "yarn.nodemanager.linux-container-executor.cgroups.hierarchy":"yarn",
    "yarn.nodemanager.container-
executor.class":"org.apache.hadoop.yarn.server.nodemanager.LinuxContainerExecutor"
  }
},{
  "Classification":"container-executor",
  "Properties":{

  },
  "Configurations":[
    {
      "Classification":"gpu",
      "Properties":{
        "module.enabled":"true"
      }
    }
  ]
}
```

```

    },
    {
      "Classification": "cgroups",
      "Properties": {
        "root": "/spark-rapids-cgroup",
        "yarn-hierarchy": "yarn"
      }
    }
  ]
}

```

## Amazon EMR 6.x

### Beispiel für eine YARN-Konfiguration für Amazon EMR 6.x

```

{
  "Classification": "yarn-site",
  "Properties": {
    "yarn.nodemanager.resource-plugins": "yarn.io/gpu",
    "yarn.resource-types": "yarn.io/gpu",
    "yarn.nodemanager.resource-plugins.gpu.allowed-gpu-devices": "auto",
    "yarn.nodemanager.resource-plugins.gpu.path-to-discovery-executables": "/usr/
bin",
    "yarn.nodemanager.linux-container-executor.cgroups.mount": "true",
    "yarn.nodemanager.linux-container-executor.cgroups.mount-path": "/sys/fs/
cgroup",
    "yarn.nodemanager.linux-container-executor.cgroups.hierarchy": "yarn",
    "yarn.nodemanager.container-
executor.class": "org.apache.hadoop.yarn.server.nodemanager.LinuxContainerExecutor"
  }
}, {
  "Classification": "container-executor",
  "Properties": {

  },
  "Configurations": [
    {
      "Classification": "gpu",
      "Properties": {
        "module.enabled": "true"
      }
    },
    {
      "Classification": "cgroups",

```

```

        "Properties":{
            "root":"/sys/fs/cgroup",
            "yarn-hierarchy":"yarn"
        }
    ]
}

```

### 3. Spark für die Verwendung von RAPIDS konfigurieren

Hier sind die erforderlichen Konfigurationen, damit Spark das RAPIDS-Plugin verwenden kann:

```

{
  "Classification":"spark-defaults",
  "Properties":{
    "spark.plugins":"com.nvidia.spark.SQLPlugin",
    "spark.executor.resource.gpu.discoveryScript":"/usr/lib/spark/scripts/gpu/getGpusResources.sh",
    "spark.executor.extraLibraryPath":"/usr/local/cuda/targets/x86_64-linux/lib:/usr/local/cuda/extras/CUPTI/lib64:/usr/local/cuda/compat/lib:/usr/local/cuda/lib:/usr/local/cuda/lib64:/usr/lib/hadoop/lib/native:/usr/lib/hadoop-lzo/lib/native:/docker/usr/lib/hadoop/lib/native:/docker/usr/lib/hadoop-lzo/lib/native"
  }
}

```

Die [XGBoost4J-Spark-Bibliothek](#) in der XGBoost-Dokumentation ist auch verfügbar, wenn das Spark-RAPIDS-Plugin auf Ihrem Cluster aktiviert ist. Sie können die folgende Konfiguration verwenden, um XGBoost in Ihren Spark-Auftrag zu integrieren:

```

{
  "Classification":"spark-defaults",
  "Properties":{
    "spark.submit.pyFiles":"/usr/lib/spark/jars/xgboost4j-spark_3.0-1.4.2-0.3.0.jar"
  }
}

```

Weitere Spark-Konfigurationen, mit denen Sie einen GPU-beschleunigten EMR-Cluster optimieren können, finden Sie im [Accelerator für Apache-Spark-Tuning-Anleitung](#) in der Nvidia.github.io-Dokumentation.

### 4. YARN Capacity Scheduler konfigurieren

DominantResourceCalculator muss so konfiguriert sein, dass GPU-Planung und -Isolierung aktiviert werden. Weitere Informationen finden Sie unter [Verwendung der GPU auf YARN](#) in der Apache-Hadoop-Dokumentation.

```
{
  "Classification": "capacity-scheduler",
  "Properties": {
    "yarn.scheduler.capacity.resource-
calculator": "org.apache.hadoop.yarn.util.resource.DominantResourceCalculator"
  }
}
```

5. Eine JSON-Datei erstellen, die all Ihre Konfigurationen enthält

Sie können eine JSON-Datei erstellen, die Ihre Konfiguration für die Verwendung des RAPIDS-Plugins für Ihren Spark-Cluster enthält. Sie stellen die Datei später bereit, wenn Sie Ihren Cluster starten.

Sie können die Datei lokal oder in S3 speichern. Weitere Informationen zur Bereitstellung von Anwendungskonfigurationen für Ihre Cluster finden Sie unter [Anwendungen konfigurieren](#).

Verwenden Sie die folgenden Beispieldateien als Vorlagen, um Ihre eigenen Konfigurationen zu erstellen.

Amazon EMR 7.x

**my-configurations.json**-Beispieldatei für Amazon EMR 7.x

```
[
  {
    "Classification": "spark",
    "Properties": {
      "enableSparkRapids": "true"
    }
  },
  {
    "Classification": "yarn-site",
    "Properties": {
      "yarn.nodemanager.resource-plugins": "yarn.io/gpu",
      "yarn.resource-types": "yarn.io/gpu",
      "yarn.nodemanager.resource-plugins.gpu.allowed-gpu-devices": "auto",
      "yarn.nodemanager.resource-plugins.gpu.path-to-discovery-executables": "/usr/bin",

```

```

        "yarn.nodemanager.linux-container-executor.cgroups.mount":"true",
        "yarn.nodemanager.linux-container-executor.cgroups.mount-path":"/spark-
rapids-cgroup",
        "yarn.nodemanager.linux-container-executor.cgroups.hierarchy":"yarn",
        "yarn.nodemanager.container-
executor.class":"org.apache.hadoop.yarn.server.nodemanager.LinuxContainerExecutor"
    }
},
{
    "Classification":"container-executor",
    "Properties":{

    },
    "Configurations":[
        {
            "Classification":"gpu",
            "Properties":{
                "module.enabled":"true"
            }
        },
        {
            "Classification":"cgroups",
            "Properties":{
                "root":"/spark-rapids-cgroup",
                "yarn-hierarchy":"yarn"
            }
        }
    ]
},
{
    "Classification":"spark-defaults",
    "Properties":{
        "spark.plugins":"com.nvidia.spark.SQLPlugin",
        "spark.executor.resource.gpu.discoveryScript":"/usr/lib/spark/scripts/
gpu/getGpusResources.sh",
        "spark.executor.extraLibraryPath":"/usr/local/cuda/targets/x86_64-
linux/lib:/usr/local/cuda/extras/CUPTI/lib64:/usr/local/cuda/compat/lib:/usr/local/
cuda/lib:/usr/local/cuda/lib64:/usr/lib/hadoop/lib/native:/usr/lib/hadoop-lzo/lib/
native:/docker/usr/lib/hadoop/lib/native:/docker/usr/lib/hadoop-lzo/lib/native",
        "spark.submit.pyFiles":"/usr/lib/spark/jars/xgboost4j-
spark_3.0-1.4.2-0.3.0.jar",
        "spark.rapids.sql.concurrentGpuTasks":"1",
        "spark.executor.resource.gpu.amount":"1",
        "spark.executor.cores":"2",

```

```

        "spark.task.cpus":"1",
        "spark.task.resource.gpu.amount":"0.5",
        "spark.rapids.memory.pinnedPool.size":"0",
        "spark.executor.memoryOverhead":"2G",
        "spark.locality.wait":"0s",
        "spark.sql.shuffle.partitions":"200",
        "spark.sql.files.maxPartitionBytes":"512m"
    }
},
{
    "Classification":"capacity-scheduler",
    "Properties":{
        "yarn.scheduler.capacity.resource-
calculator":"org.apache.hadoop.yarn.util.resource.DominantResourceCalculator"
    }
}
]

```

## Amazon EMR 6.x

### my-configurations.json-Beispieldatei für Amazon EMR 6.x

```

[
  {
    "Classification":"spark",
    "Properties":{
      "enableSparkRapids":"true"
    }
  },
  {
    "Classification":"yarn-site",
    "Properties":{
      "yarn.nodemanager.resource-plugins":"yarn.io/gpu",
      "yarn.resource-types":"yarn.io/gpu",
      "yarn.nodemanager.resource-plugins.gpu.allowed-gpu-devices":"auto",
      "yarn.nodemanager.resource-plugins.gpu.path-to-discovery-executables":"/
usr/bin",
      "yarn.nodemanager.linux-container-executor.cgroups.mount":"true",
      "yarn.nodemanager.linux-container-executor.cgroups.mount-path":"/sys/fs/
cgroup",
      "yarn.nodemanager.linux-container-executor.cgroups.hierarchy":"yarn",
      "yarn.nodemanager.container-
executor.class":"org.apache.hadoop.yarn.server.nodemanager.LinuxContainerExecutor"
    }
  }
]

```



```
},
{
  "Classification": "container-executor",
  "Properties": {
  },
  "Configurations": [
    {
      "Classification": "gpu",
      "Properties": {
        "module.enabled": "true"
      }
    },
    {
      "Classification": "cgroups",
      "Properties": {
        "root": "/sys/fs/cgroup",
        "yarn-hierarchy": "yarn"
      }
    }
  ]
},
{
  "Classification": "spark-defaults",
  "Properties": {
    "spark.plugins": "com.nvidia.spark.SQLPlugin",
    "spark.executor.resource.gpu.discoveryScript": "/usr/lib/spark/scripts/gpu/getGpusResources.sh",
    "spark.executor.extraLibraryPath": "/usr/local/cuda/targets/x86_64-linux/lib:/usr/local/cuda/extras/CUPTI/lib64:/usr/local/cuda/compat/lib:/usr/local/cuda/lib:/usr/local/cuda/lib64:/usr/lib/hadoop/lib/native:/usr/lib/hadoop-lzo/lib/native:/docker/usr/lib/hadoop/lib/native:/docker/usr/lib/hadoop-lzo/lib/native",
    "spark.submit.pyFiles": "/usr/lib/spark/jars/xgboost4j-spark_3.0-1.4.2-0.3.0.jar",
    "spark.rapids.sql.concurrentGpuTasks": "1",
    "spark.executor.resource.gpu.amount": "1",
    "spark.executor.cores": "2",
    "spark.task.cpus": "1",
    "spark.task.resource.gpu.amount": "0.5",
    "spark.rapids.memory.pinnedPool.size": "0",
    "spark.executor.memoryOverhead": "2G",
    "spark.locality.wait": "0s",
    "spark.sql.shuffle.partitions": "200",
    "spark.sql.files.maxPartitionBytes": "512m"
  }
}
```

```
    }  
  },  
  {  
    "Classification": "capacity-scheduler",  
    "Properties": {  
      "yarn.scheduler.capacity.resource-  
calculator": "org.apache.hadoop.yarn.util.resource.DominantResourceCalculator"  
    }  
  }  
]  
]
```

## Fügen Sie eine Bootstrap-Aktion für Ihren Cluster hinzu

Weitere Informationen zur Bereitstellung von Bootstrap-Aktionsskripten bei der Erstellung Ihres Clusters finden Sie unter [Grundlagen der Bootstrap-Aktion](#) im Verwaltungshandbuch für Amazon EMR.

Die folgenden Beispielskripts zeigen, wie eine Bootstrap-Aktionsdatei für Amazon EMR 6.x und 7.x erstellt wird:

### Amazon EMR 7.x

#### **my-bootstrap-action.sh**-Beispieldatei für Amazon EMR 7.x

Um YARN zur Verwaltung von GPU-Ressourcen mit Amazon-EMR-7.x-Versionen zu verwenden, müssen Sie die CGroup v1 manuell auf Ihrem Cluster bereitstellen. Sie können dies mit einem Bootstrap-Aktionsskript tun, wie in diesem Beispiel gezeigt.

```
#!/bin/bash  
set -ex  
  
sudo mkdir -p /spark-rapids-cgroup/devices  
sudo mount -t cgroup -o devices cgroupv1-devices /spark-rapids-cgroup/devices  
sudo chmod a+rx -R /spark-rapids-cgroup
```

### Amazon EMR 6.x

#### **my-bootstrap-action.sh**-Beispieldatei für Amazon EMR 6.x

Für Amazon-EMR-6.x-Versionen müssen Sie CGroup-Berechtigungen für YARN in Ihrem Cluster öffnen. Sie können dies mit einem Bootstrap-Aktionsskript tun, wie in diesem Beispiel gezeigt.

```
#!/bin/bash
set -ex

sudo chmod a+rxw -R /sys/fs/cgroup/cpu,cpuacct
sudo chmod a+rxw -R /sys/fs/cgroup/devices
```

## Ihren Cluster starten

Der letzte Schritt besteht darin, Ihren Cluster mit den oben genannten Cluster-Konfigurationen zu starten. Hier ist ein Beispielbefehl zum Starten eines Clusters über die Amazon-EMR-CLI:

```
aws emr create-cluster \
--release-label emr-7.1.0 \
--applications Name=Hadoop Name=Spark \
--service-role EMR_DefaultRole_V2 \
--ec2-attributes KeyName=my-key-pair,InstanceProfile=EMR_EC2_DefaultRole \
--instance-groups InstanceGroupType=MASTER,InstanceCount=1,InstanceType=m4.4xlarge \
                    InstanceGroupType=CORE,InstanceCount=1,InstanceType=g4dn.2xlarge \
                    InstanceGroupType=TASK,InstanceCount=1,InstanceType=g4dn.2xlarge \
--configurations file:///my-configurations.json \
--bootstrap-actions Name='My Spark Rapids Bootstrap action',Path=s3://my-bucket/my-
bootstrap-action.sh
```

## Zugriff auf die Spark-Shell

Die Spark-Shell basiert auf dem Scala REPL (Read-Eval-Print-Loop). Damit können Sie Spark-Programme interaktiv erstellen und Arbeit an das Framework senden. Sie können auf die Spark-Shell zugreifen, indem Sie eine Verbindung mit dem Primärknoten über SSH herstellen und `spark-shell` aufrufen. Weitere Informationen zum Herstellen einer Verbindung mit dem Primärknoten finden Sie unter [Verbinden mit dem Primärknoten über SSH](#) im Verwaltungshandbuch für Amazon EMR. In den folgenden Beispielen werden Apache-HTTP-Serverzugriffsprotokolle in Amazon S3 verwendet.

### Note

Der Bucket in diesen Beispielen ist für Clients mit Zugriff auf USA Ost (Nord-Virginia) verfügbar.

Standardmäßig erstellt die Spark-Shell ihr eigenes [SparkContext](#) Objekt namens `sc`. Sie können diesen Kontext verwenden, wenn er in der REPL erforderlich ist. `sqlContext` ist auch in der Shell verfügbar und es ist ein [HiveContext](#).

**Example** Verwenden Sie die Spark-Shell, um die Vorkommen einer Zeichenkette in einer in Amazon S3 gespeicherten Datei zu zählen

Dieses Beispiel verwendet `sc` zum Lesen einer Textdatei in Amazon S3.

```
scala> sc
res0: org.apache.spark.SparkContext = org.apache.spark.SparkContext@404721db

scala> val textFile = sc.textFile("s3://elasticmapreduce/samples/hive-ads/tables/
impressions/dt=2009-04-13-08-05/ec2-0-51-75-39.amazon.com-2009-04-13-08-05.log")
```

Spark erstellt die Datei „textFile“ und die zugehörige [-Datenstruktur](#). Im Beispiel wird dann die Anzahl der Zeilen in der Protokolldatei mit der Zeichenfolge "cartoonnetwork.com" gezählt:

```
scala> val linesWithCartoonNetwork = textFile.filter(line =>
  line.contains("cartoonnetwork.com")).count()
linesWithCartoonNetwork: org.apache.spark.rdd.RDD[String] = MapPartitionsRDD[2] at
  filter at <console>:23
<snip>
<Spark program runs>
scala> linesWithCartoonNetwork
res2: Long = 9
```

**Example** Verwenden Sie die Python-basierte Spark-Shell, um die Vorkommen einer Zeichenkette in einer in Amazon S3 gespeicherten Datei zu zählen

Spark umfasst auch eine auf Python basierende Shell, `pyspark`, mit der Sie Prototypen von in Python geschriebenen Spark-Programmen entwickeln können. Genau wie `beispark-shell`, `pyspark` auf dem Primärknoten aufrufen; es hat auch dasselbe [SparkContext](#) Objekt.

```
>>> sc
<pyspark.context.SparkContext object at 0x7fe7e659fa50>
>>> textfile = sc.textFile("s3://elasticmapreduce/samples/hive-ads/tables/impressions/
dt=2009-04-13-08-05/ec2-0-51-75-39.amazon.com-2009-04-13-08-05.log")
```

Spark erstellt die Datei „textFile“ und die zugehörige [-Datenstruktur](#). Im Beispiel wird dann die Anzahl der Zeilen in der Protokolldatei mit der Zeichenfolge "cartoonnetwork.com" gezählt.

```
>>> linesWithCartoonNetwork = textfile.filter(lambda line: "cartoonnetwork.com" in
line).count()
15/06/04 17:12:22 INFO lzo.GPLNativeCodeLoader: Loaded native gpl library from the
embedded binaries
15/06/04 17:12:22 INFO lzo.LzoCodec: Successfully loaded & initialized native-lzo
library [hadoop-lzo rev EXAMPLE]
15/06/04 17:12:23 INFO fs.EmrFileSystem: Consistency disabled, using
com.amazon.ws.emr.hadoop.fs.s3n.S3NativeFileSystem as filesystem implementation
<snip>
<Spark program continues>
>>> linesWithCartoonNetwork
9
```

## Verwenden Sie Amazon SageMaker Spark für maschinelles Lernen

Bei der Verwendung von Amazon-EMR-Version 5.11.0 und höher wird die `aws-sagemaker-spark-sdk`-Komponente zusammen mit Spark installiert. Diese Komponente installiert Amazon SageMaker Spark und die zugehörigen Abhängigkeiten für die Spark-Integration mit [Amazon SageMaker](#). Sie können Amazon SageMaker Spark verwenden, um Spark-Pipelines für maschinelles Lernen (ML) mithilfe von SageMaker Amazon-Stufen zu erstellen. Weitere Informationen finden Sie in der [Amazon SageMaker Spark-README-Datei](#) zu Apache Spark GitHub und zur [Verwendung von Apache Spark mit Amazon SageMaker](#) im Amazon SageMaker Developer Guide.

## Eine Spark-Anwendung schreiben

[Spark](#)-Anwendungen können in Scala, Java und Python geschrieben werden. In der Dokumentation zu Apache Spark finden Sie im Thema [Spark-Beispiele](#) einige Beispiele für Spark-Anwendungen. Das Beispiel zum Schätzen von Pi, "Estimating Pi", ist unten in den drei nativ unterstützten Anwendungen dargestellt. Vollständige Beispiele finden Sie auch unter `$SPARK_HOME/examples` und unter [GitHub](#). Weitere Informationen zum Erstellen von JAR-Dateien für Spark finden Sie im Thema [Schnellstart](#) in der Apache-Spark-Dokumentation.

## Scala

Um Scala-Kompatibilitätsprobleme zu vermeiden, empfehlen wir, Spark-Abhängigkeiten für die richtige Scala-Version zu verwenden, wenn Sie eine Spark-Anwendung für einen Amazon-EMR-Cluster kompilieren. Welche Scala-Version Sie verwenden sollten, hängt von der Version von Spark ab, die auf Ihrem Cluster installiert ist. Amazon-EMR-Version 5.30.1 verwendet beispielsweise Spark

2.4.5, das mit Scala 2.11 erstellt wurde. Wenn Ihr Cluster Amazon-EMR-Version 5.30.1 verwendet, verwenden Sie Spark-Abhängigkeiten für Scala 2.11. Weitere Informationen zu den von Spark verwendeten Scala-Versionen finden Sie in der [Apache-Spark-Dokumentation](#).

```
package org.apache.spark.examples
import scala.math.random
import org.apache.spark._

/** Computes an approximation to pi */
object SparkPi {
  def main(args: Array[String]) {
    val conf = new SparkConf().setAppName("Spark Pi")
    val spark = new SparkContext(conf)
    val slices = if (args.length > 0) args(0).toInt else 2
    val n = math.min(100000L * slices, Int.MaxValue).toInt // avoid overflow
    val count = spark.parallelize(1 until n, slices).map { i =>
      val x = random * 2 - 1
      val y = random * 2 - 1
      if (x*x + y*y < 1) 1 else 0
    }.reduce(_ + _)
    println("Pi is roughly " + 4.0 * count / n)
    spark.stop()
  }
}
```

## Java

```
package org.apache.spark.examples;

import org.apache.spark.SparkConf;
import org.apache.spark.api.java.JavaRDD;
import org.apache.spark.api.java.JavaSparkContext;
import org.apache.spark.api.java.function.Function;
import org.apache.spark.api.java.function.Function2;

import java.util.ArrayList;
import java.util.List;

/**
 * Computes an approximation to pi
 * Usage: JavaSparkPi [slices]
 */
```

```
public final class JavaSparkPi {

    public static void main(String[] args) throws Exception {
        SparkConf sparkConf = new SparkConf().setAppName("JavaSparkPi");
        JavaSparkContext jsc = new JavaSparkContext(sparkConf);

        int slices = (args.length == 1) ? Integer.parseInt(args[0]) : 2;
        int n = 100000 * slices;
        List<Integer> l = new ArrayList<Integer>(n);
        for (int i = 0; i < n; i++) {
            l.add(i);
        }

        JavaRDD<Integer> dataSet = jsc.parallelize(l, slices);

        int count = dataSet.map(new Function<Integer, Integer>() {
            @Override
            public Integer call(Integer integer) {
                double x = Math.random() * 2 - 1;
                double y = Math.random() * 2 - 1;
                return (x * x + y * y < 1) ? 1 : 0;
            }
        }).reduce(new Function2<Integer, Integer, Integer>() {
            @Override
            public Integer call(Integer integer, Integer integer2) {
                return integer + integer2;
            }
        });

        System.out.println("Pi is roughly " + 4.0 * count / n);

        jsc.stop();
    }
}
```

## Python

```
import argparse
import logging
from operator import add
from random import random
```

```
from pyspark.sql import SparkSession

logger = logging.getLogger(__name__)
logging.basicConfig(level=logging.INFO, format="%(levelname)s: %(message)s")

def calculate_pi(partitions, output_uri):
    """
    Calculates pi by testing a large number of random numbers against a unit circle
    inscribed inside a square. The trials are partitioned so they can be run in
    parallel on cluster instances.

    :param partitions: The number of partitions to use for the calculation.
    :param output_uri: The URI where the output is written, typically an Amazon S3
                       bucket, such as 's3://example-bucket/pi-calc'.
    """

    def calculate_hit(_):
        x = random() * 2 - 1
        y = random() * 2 - 1
        return 1 if x**2 + y**2 < 1 else 0

    tries = 100000 * partitions
    logger.info(
        "Calculating pi with a total of %s tries in %s partitions.", tries, partitions
    )
    with SparkSession.builder.appName("My PyPi").getOrCreate() as spark:
        hits = (
            spark.sparkContext.parallelize(range(tries), partitions)
            .map(calculate_hit)
            .reduce(add)
        )
        pi = 4.0 * hits / tries
        logger.info("%s tries and %s hits gives pi estimate of %s.", tries, hits, pi)
        if output_uri is not None:
            df = spark.createDataFrame([(tries, hits, pi)], ["tries", "hits", "pi"])
            df.write.mode("overwrite").json(output_uri)

if __name__ == "__main__":
    parser = argparse.ArgumentParser()
    parser.add_argument(
        "--partitions",
        default=2,
```



```
        type=int,
        help="The number of parallel partitions to use when calculating pi.",
    )
    parser.add_argument(
        "--output_uri", help="The URI where output is saved, typically an S3 bucket."
    )
    args = parser.parse_args()

    calculate_pi(args.partitions, args.output_uri)
```

## Verbessern der Spark Leistung mit Amazon S3

Amazon EMR bietet Features, mit denen die Leistung optimiert werden kann, wenn Spark für Abfragen und Lese- bzw. Schreiboperationen über Daten in Amazon S3 verwendet wird.

[S3 Select](#) kann in einigen Anwendungen die Abfrageleistung bei CSV- und JSON-Dateien verbessern, indem die Verarbeitung an Amazon S3 ausgelagert wird.

Der EMRFS S3-optimierte Committer ist eine Alternative zu der [OutputCommitter](#)-Klasse, die die Funktion für mehrteilige Uploads von EMRFS verwendet, um die Leistung beim Schreiben von Parquet-Dateien auf Amazon S3 mithilfe von Spark SQL, und Datasets zu verbessern. DataFrames

Themen

- [S3 Select mit Spark zur Verbesserung der Leistung bei Abfragen verwenden](#)
- [EMRFS-S3-optimierte Committer verwenden](#)
- [Verwenden Sie das EMRFS-S3-optimierte Commit-Protokoll](#)
- [Amazon-S3-Anforderungen mit EMRFS wiederholen](#)

## S3 Select mit Spark zur Verbesserung der Leistung bei Abfragen verwenden

Mit Amazon-EMR-Version 5.17.0 und höher können Sie [S3 Select](#) mit Spark auf Amazon EMR verwenden. S3 Select ermöglicht es Anwendungen, nur eine Teilmenge von Daten aus einem Objekt abzurufen. Bei Amazon EMR wird die numerische Arbeit zur Filterung großer Datensätze von dem Cluster an Amazon S3 ausgelagert. Dies kann die Leistung in einigen Anwendungen verbessern und reduziert die Menge der zwischen Amazon EMR und Amazon S3 übertragenen Daten.

S3 Select unterstützt CSV und JSON-Dateien, das Datenformat wird unter Verwendung von `s3selectCSV` und `s3selectJSON` angegeben. Weitere Informationen und Beispiele finden Sie unter [S3 Select in Ihrem Code angeben](#).

## Ist S3 Select das Richtige für meine Anwendung?

Wir empfehlen, dass Sie Benchmark-Tests für Ihre Anwendungen im Vergleich mit und ohne S3 Select durchführen, um zu sehen, ob es für Ihre Anwendung geeignet sein könnte.

Verwenden Sie die folgenden Richtlinien, um zu bestimmen, ob Ihre Anwendung ein Kandidat für die Verwendung von S3 ist:

- Ihre Abfrage filtert mehr als die Hälfte des ursprünglichen Datensatzes.
- Ihre Netzwerkverbindung zwischen Amazon S3 und dem Amazon-EMR-Cluster hat eine gute Übertragungsgeschwindigkeit und verfügbare Bandbreite. Amazon S3 komprimiert keine HTTP-Antworten, sodass die Antwortgröße bei komprimierten Eingabedateien wahrscheinlich zunimmt.

## Überlegungen und Einschränkungen

- Die Amazon-S3-serverseitige Verschlüsselung mit vom Kunden bereitgestellten Verschlüsselungsschlüsseln (SSE-C) und die clientseitige Verschlüsselung werden nicht unterstützt.
- Die Eigenschaft `AllowQuotedRecordDelimiters` wird nicht unterstützt. Wenn diese Eigenschaft angegeben ist, schlägt die Abfrage fehl.
- Nur CSV- und JSON-Dateien im UTF-8-Format werden unterstützt. Es werden keine CSV-Datensätze unterstützt, die sich über mehrere Zeilen erstrecken.
- Es werden nur unkomprimierte oder gzip-Dateien unterstützt.
- Die Optionen für die Datenformate CSV und JSON in Spark (z. B. `nanValue`, `positiveInf`, `negativeInf`) sowie Optionen zum Umgang mit beschädigten Datensätzen (z. B. der `failfast`-Modus zur frühen Terminierung und der `dropmalformed`-Modus zum Überspringen von Datensätzen im falschen Format) werden nicht unterstützt.
- Die Verwendung von Kommas als Tausender-Trennzeichen wird bei Dezimalzahlen nicht unterstützt. Beispielsweise ist `10,000` ungültig, `10000` wird unterstützt.
- Kommentarzeichen auf der letzten Zeile werden nicht unterstützt.
- Leere Zeilen am Ende einer Datei werden nicht verarbeitet.
- Die folgenden Filter werden nicht an Amazon S3 ausgelagert:

- Aggregatfunktionen wie COUNT() und SUM().
- Filter, die ein Attribut über einen CAST()-Aufruf konvertieren. z. B. CAST(stringColumn as INT) = 1.
- Filter mit einem Attribut, das ein Objekt ist oder einem zusammengesetzten Attribut. z. B. intArray[1] = 1, objectColumn.objectNumber = 1.
- Filter, deren Wert nicht literal ist. Beispiel: intColumn1 = intColumn2
- Es werden nur die [von S3 Select unterstützten Datentypen](#) unterstützt, und es gelten die dokumentierten Einschränkungen.

## S3 Select in Ihrem Code angeben

Die folgenden Beispiele zeigen, wie S3 Select für CSV mithilfe von Scala, SQL, R und PySpark angegeben wird. Auf die gleiche Weise können Sie mit S3 Select JSON als Datenformat festlegen. Eine Liste der Optionen, der Standardwerte und der Einschränkungen finden Sie unter [Optionen](#).

### PySpark

```
spark
  .read
  .format("s3selectCSV") // "s3selectJson" for Json
  .schema(...) // optional, but recommended
  .options(...) // optional
  .load("s3://path/to/my/datafiles")
```

### R

```
read.df("s3://path/to/my/datafiles", "s3selectCSV", schema, header = "true",
  delimiter = "\t")
```

### Scala

```
spark
  .read
  .format("s3selectCSV") // "s3selectJson" for Json
  .schema(...) // optional, but recommended
  .options(...) // optional. Examples:
  // .options(Map("quote" -> "\"", "header" -> "true")) or
  // .option("quote", "\"").option("header", "true")
```

```
.load("s3://path/to/my/datafiles")
```

## SQL

```
CREATE TEMPORARY VIEW MyView (number INT, name STRING) USING s3selectCSV OPTIONS
(path "s3://path/to/my/datafiles", header "true", delimiter "\t")
```

## Optionen

Die folgenden Optionen sind verfügbar, wenn `s3selectCSV` und `s3selectJSON` verwendet wird. Wenn nichts angegeben wird, werden die Standardwerte verwendet.

### Optionen für S3selectCSV

| Option                   | Standard | Verwendung                                                                                                                                                                             |
|--------------------------|----------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <code>compression</code> | "none"   | Gibt an, ob Komprimierung verwendet wird. "gzip" wird neben "none" als einzige Einstellung unterstützt.                                                                                |
| <code>delimiter</code>   | ","      | Gibt den Feldbegrenzer an.                                                                                                                                                             |
| <code>quote</code>       | '\"'     | Gibt das Zeichen an, das als Anführungszeichen verwendet wird. Die Angabe einer leeren Zeichenfolge wird nicht unterstützt und führt zu einem XML-Fehler aufgrund fehlerhaftem Format. |
| <code>escape</code>      | '\\'     | Gibt das Zeichen an, das als Escape-Zeichen verwendet wird.                                                                                                                            |
| <code>header</code>      | "false"  | "false" gibt an, dass es keine Kopfzeile gibt. "true" gibt an, dass die erste Zeile                                                                                                    |

| Option    | Standard | Verwendung                                                                                                                                                                                   |
|-----------|----------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|           |          | die Kopfzeile ist. Es werden ausschließlich Kopfzeilen in der ersten Zeile unterstützt, und leere Zeilen vor der Kopfzeile werden nicht unterstützt.                                         |
| Kommentar | "#"      | Gibt das Zeichen an, das als Kommentarzeichen verwendet wird. Der Kommentarindikator kann nicht deaktiviert werden. Dies bedeutet, dass der Wert <code>\u0000</code> nicht unterstützt wird. |
| nullValue | ""       |                                                                                                                                                                                              |

### Optionen für S3selectJSON

| Option      | Standard | Verwendung                                                                                                                                                                                                                                                            |
|-------------|----------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| compression | "none"   | Gibt an, ob Komprimierung verwendet wird. "gzip" wird neben "none" als einzige Einstellung unterstützt.                                                                                                                                                               |
| multiline   | „false“  | "false" gibt an, dass die JSON-Daten in S3 Select im LINES-Format vorliegt. Dies bedeutet, dass jede Zeile in den Eingabedaten genau ein JSON-Objekt enthält. "true" gibt an, dass die JSON-Daten in S3 Select im DOCUMENT-Format vorliegen. Dies bedeutet, dass sich |

| Option | Standard | Verwendung                                                               |
|--------|----------|--------------------------------------------------------------------------|
|        |          | ein JSON-Objekt in den Eingabedaten über mehrere Zeilen erstrecken kann. |

## EMRFS-S3-optimierte Committer verwenden

Der EMRFS S3-optimierte Committer ist eine alternative [OutputCommitter](#) Implementierung, die für das Schreiben von Dateien in Amazon S3 bei Verwendung von EMRFS optimiert ist. Der für EMRFS S3 optimierte Comitter verbessert die Anwendungsleistung durch Umgehung der Operationen zum Auflisten und Umbenennen, die während der Commit-Phasen von Aufträgen und Aufgaben in Amazon S3 durchgeführt werden. Die Committer-Klasse ist verfügbar für Amazon-EMR-Version 5.19.0 und höher und ist bei Amazon EMR 5.20.0 und höher standardmäßig aktiviert. Der Committer wird für Spark-Jobs verwendet, die Spark SQL oder Datasets verwenden. DataFrames Ab Amazon EMR 6.4.0 kann dieser Committer für alle gängigen Formate verwendet werden, einschließlich Parquet, ORC und textbasierte Formate (einschließlich CSV und JSON). Für Versionen vor Amazon EMR 6.4.0 wird nur das Parquet-Format unterstützt. Es gibt Situationen, in denen der Committer nicht verwendet wird. Weitere Informationen finden Sie unter [Anforderungen für den S3-optimierten EMRFS-Committer](#).

### Themen

- [Anforderungen für den S3-optimierten EMRFS-Committer](#)
- [Der S3-optimierte EMRFS-Committer und mehrteilige Uploads](#)
- [Überlegungen zur Auftragsoptimierung](#)
- [Aktivieren Sie den S3-optimierten Committer für Amazon EMR 5.19.0](#)


## Anforderungen für den S3-optimierten EMRFS-Committer

Der S3-optimierte EMRFS-Committer wird verwendet, wenn die folgenden Bedingungen erfüllt sind:

- Sie führen Spark-Jobs aus, die Spark SQL oder Datasets verwenden DataFrames, um Dateien in Amazon S3 zu schreiben. Ab Amazon EMR 6.4.0 kann dieser Committer für alle gängigen Formate verwendet werden, einschließlich Parquet, ORC und textbasierte Formate (einschließlich CSV und JSON). Für Versionen vor Amazon EMR 6.4.0 wird nur das Parquet-Format unterstützt.

- In Amazon EMR sind mehrteilige Uploads aktiviert. Dies ist die Standardeinstellung. Weitere Informationen finden Sie unter [Der S3-optimierte EMRFS-Committer und mehrteilige Uploads](#).
- Die integrierte Dateiformatunterstützung von Spark wird verwendet. Die integrierte Dateiformatunterstützung wird unter folgenden Umständen verwendet:
  - Bei Hive-Metastore-Tabellen, wenn `spark.sql.hive.convertMetastoreParquet` für Parquet-Tabellen auf `true` oder `spark.sql.hive.convertMetastoreOrc` für Orc-Tabellen mit Amazon EMR 6.4.0 oder höher auf `true` eingestellt. Dies sind die Standardeinstellungen.
  - Wenn Aufträge in Datenquellen oder Tabellen im Dateiformat schreiben, wird beispielsweise die Zieltabelle mit der `USING parquet` Klausel erstellt.
  - Wenn Aufträge in nicht partitionierte Hive-Metastore-Parquet-Tabellen schreiben. Eine bekannte Einschränkung besteht darin, dass die in Spark integrierte Parquet-Unterstützung keine partitionierten Hive-Tabellen unterstützt. Weitere Informationen finden Sie unter [Hive Metastore Parquet-Tabellenkonvertierung](#) im Apache Spark SQL DataFrames and Datasets Guide.
- Spark-Job-Operationen, die beispielsweise `${table_location}/k1=v1/k2=v2/` in einen Standardspeicherort für Partitionen schreiben, verwenden den Committer. Der Committer wird nicht verwendet, wenn ein Job-Vorgang in einen benutzerdefinierten Partitionsspeicherort schreibt, z. B. wenn mit dem `ALTER TABLE SQL`-Befehl ein benutzerdefinierter Partitionsspeicherort festgelegt wurde.
- Für Spark müssen die folgenden Werte verwendet werden:
  - Der Eigenschaft `spark.sql.parquet.fs.optimized.committer.optimization-enabled` muss auf `true` eingestellt sein. Dies ist die Standardeinstellung bei Amazon EMR 5.20.0 und höher. Bei Amazon EMR 5.19.0 lautet der Standardwert `false`. Weitere Informationen zum Konfigurieren dieses Werts finden Sie unter [Aktivieren Sie den S3-optimierten Committer für Amazon EMR 5.19.0](#).
  - Beim Schreiben in nicht partitionierte Hive-Metastore-Tabellen werden nur die Dateiformate Parquet und Orc unterstützt. `spark.sql.hive.convertMetastoreParquet` muss auf `true` gesetzt sein, wenn in nicht partitionierte Paquet Hive-Metastore-Tabellen geschrieben werden. `spark.sql.hive.convertMetastoreOrc` muss auf `true` gesetzt werden, wenn in nicht partitionierte Orc-Hive-Metastore-Tabellen geschrieben wird. Dies sind die Standardeinstellungen.
  - muss `spark.sql.parquet.output.committer.class` auf `com.amazon.emr.committer.EmrOptimizedSparkSqlParquetOutputCommitter` festgelegt sein. Dies ist die Standardeinstellung.

- `spark.sql.sources.commitProtocolClass` muss auf `org.apache.spark.sql.execution.datasources.SQLEmrOptimizedCommitProtocol` oder `org.apache.spark.sql.execution.datasources.SQLHadoopMapReduceCommitProtocol` eingestellt sein.  
`org.apache.spark.sql.execution.datasources.SQLEmrOptimizedCommitProtocol` ist die Standardeinstellung für die Amazon-EMR-5.x-Serie, Version 5.30.0 und höher, und für die Amazon-EMR-6.x-Serie, Version 6.2.0 und höher.  
`org.apache.spark.sql.execution.datasources.SQLHadoopMapReduceCommitProtocol` ist die Standardeinstellung für frühere Amazon-EMR-Versionen.
- Wenn Spark-Aufträge partitionierte Parquet-Datasets durch dynamische Partitionsspalten überschreiben, dann müssen die Schreiboption `partitionOverwriteMode` und `spark.sql.sources.partitionOverwriteMode` auf `static` eingestellt sein. Dies ist die Standardeinstellung.

 Note

Die Schreiboption `partitionOverwriteMode` wurde mit Spark 2.4.0 eingeführt. Legen Sie für Spark-Version 2.3.2, die in Amazon-EMR-Version 5.19.0 enthalten ist, die Eigenschaft `spark.sql.sources.partitionOverwriteMode` fest.

Fälle, in denen der S3-optimierte EMRFS-Committer nicht verwendet wird

Im Allgemeinen wird der S3-optimierte Committer in den folgenden Situationen nicht verwendet.

| Situation                                             | Warum der Committer nicht verwendet wird                                                |
|-------------------------------------------------------|-----------------------------------------------------------------------------------------|
| Wenn Sie in HDFS schreiben                            | Der Committer unterstützt nur das Schreiben in Amazon S3 mit EMRFS.                     |
| Bei Verwendung des S3A-Dateisystems                   | Der Committer unterstützt nur EMRFS.                                                    |
| Wenn Sie die RDD-API von Spark verwenden<br>MapReduce | Der Committer unterstützt nur die Verwendung von SparkSQL- oder DataFrame Dataset-APIs. |



Die folgenden Scala-Beispiele veranschaulichen einige zusätzliche Situationen, die verhindern, dass der EMRFS-S3-optimierte Committer vollständig (erstes Beispiel) oder teilweise (zweites Beispiel) verwendet werden kann.

### Example – Dynamischer Partitionsüberschreibmodus

Das folgende Scala-Beispiel weist Spark an, einen anderen Commit-Algorithmus zu verwenden, wodurch die Verwendung des für EMRFS S3 optimierten Committers insgesamt verhindert wird. Der Code legt die Eigenschaft `partitionOverwriteMode` auf `dynamic` fest, sodass nur die Partitionen überschrieben werden, auf die Sie Daten schreiben. Anschließend werden dynamische Partitionsspalten durch `partitionBy` angegeben und der Schreibmodus ist auf `overwrite` eingestellt.

```
val dataset = spark.range(0, 10)
  .withColumn("dt", expr("date_sub(current_date(), id)"))

dataset.write.mode("overwrite")
  .option("partitionOverwriteMode", "dynamic")
  .partitionBy("dt")
  .parquet("s3://EXAMPLE-DOC-BUCKET/output")
```

Sie müssen alle drei Einstellungen konfigurieren, um die Verwendung des S3-optimierten Committers zu vermeiden. Wenn Sie dies tun, führt Spark einen anderen Commit-Algorithmus aus, der im Commit-Protokoll von Spark angegeben ist. Für Amazon-EMR-5.x-Versionen vor 5.30.0 und für Amazon-EMR-6.x-Versionen vor 6.2.0 verwendet das Commit-Protokoll das Staging-Verzeichnis von Spark, ein temporäres Verzeichnis, das unter dem Ausgabespeicherort erstellt wird, der mit `.spark-staging` beginnt. Der Algorithmus benennt Partitionsverzeichnisse nacheinander um, was sich negativ auf die Leistung auswirken kann. Weitere Informationen zu den Amazon-EMR-Versionen 5.30.0 und höher sowie 6.2.0 und höher finden Sie unter [Verwenden Sie das EMRFS-S3-optimierte Commit-Protokoll](#).

Der Algorithmus in Spark 2.4.0 folgt diesen Schritten:

1. Aufgabenversuche schreiben ihre Ausgabe in Partitionsverzeichnisse unterhalb des Staging-Verzeichnisses von Spark, beispielsweise `${outputLocation}/spark-staging- $\{jobID\}/k1=v1/k2=v2/$` .
2. Für jede geschriebene Partition verfolgt der Aufgabenversuch die relativen Partitionspfade, z. B. `k1=v1/k2=v2`.

3. Wenn eine Aufgabe erfolgreich abgeschlossen wurde, stellt sie dem Treiber alle zugehörigen Partitionspfade bereit, die von ihr nachverfolgt wurden.
4. Nachdem alle Aufgaben abgeschlossen wurden, sammelt die Auftrags-Commit-Phase alle die Partitionsverzeichnisse, die von erfolgreichen Aufgabenversuchen unter dem Staging-Verzeichnis von Spark geschrieben wurden. Spark benennt jedes dieser Verzeichnisse mithilfe von Verzeichnisstruktur-Umbenennungsoperationen sequenziell bis zu ihrem endgültigen Ausgabespeicherort um.
5. Das Staging-Verzeichnis wird gelöscht, bevor der Auftrags-Commit-Phase abgeschlossen ist.

### Example – Benutzerdefinierter Partitionsspeicherort

In diesem Beispiel wird der Scala-Code in zwei Partitionen eingefügt. Eine Partition verfügt über einen benutzerdefinierten Partitionsspeicherort. Die andere Partition verwendet den Standard-Partitionsspeicherort. Der S3-optimierte EMRFS-Committer wird nur zum Schreiben der Aufgabenausgabe in die Partition genutzt, die den Standard-Partitionsspeicherort verwendet.

```
val table = "dataset"
val location = "s3://bucket/table"

spark.sql(s"""
  CREATE TABLE $table (id bigint, dt date)
  USING PARQUET PARTITIONED BY (dt)
  LOCATION '$location'
  """)

// Add a partition using a custom location
val customPartitionLocation = "s3://bucket/custom"
spark.sql(s"""
  ALTER TABLE $table ADD PARTITION (dt='2019-01-28')
  LOCATION '$customPartitionLocation'
  """)

// Add another partition using default location
spark.sql(s"ALTER TABLE $table ADD PARTITION (dt='2019-01-29')")

def asDate(text: String) = lit(text).cast("date")

spark.range(0, 10)
  .withColumn("dt",
```

```
when($"id" > 4, asDate("2019-01-28")).otherwise(asDate("2019-01-29"))  
.write.insertInto(table)
```

Der Scala-Code erstellt die folgenden Amazon-S3-Objekte:

```
custom/part-00001-035a2a9c-4a09-4917-8819-e77134342402.c000.snappy.parquet  
custom_${folder$}  
table/_SUCCESS  
table/dt=2019-01-29/part-00000-035a2a9c-4a09-4917-8819-e77134342402.c000.snappy.parquet  
table/dt=2019-01-29_${folder$}  
table_${folder$}
```

Beim Schreiben in Partitionen an benutzerdefinierten Speicherorten verwendet Spark einen Commit-Algorithmus, ähnlich wie im vorherigen Beispiel. Dies wird im Folgenden beschrieben. Genau wie bei dem vorherigen Beispiel führt der Algorithmus zu sequenziellen Umbenennungen, wodurch die Leistung beeinträchtigt werden kann.

1. Beim Schreiben von Ausgabe in eine Partition an einem benutzerdefinierten Speicherort werden Aufgaben in eine Datei unter dem Staging-Verzeichnis von Spark geschrieben, das unter dem endgültigen Ausgabespeicherort erstellt wird. Der Name der Datei enthält eine zufällige UUID zum Schutz vor Datei-Kollisionen. Der Aufgabe-Versuch verfolgt jede Datei zusammen mit dem gewünschten endgültigen Ausgabepfad nach.
2. Wenn eine Aufgabe erfolgreich abgeschlossen wird, stellt sie dem Treiber die Dateien und die für sie gewünschten endgültigen Ausgabepfade bereit.
3. Nachdem alle Aufgaben beendet wurden, benennt die Auftrags-Commit-Phase alle Dateien, die für Partitionen an benutzerdefinierten Speicherorten geschrieben wurden, sequentiell in ihre endgültigen Ausgabepfade um.
4. Das Staging-Verzeichnis wird gelöscht, bevor der Auftrags-Commit-Phase abgeschlossen ist.

## Der S3-optimierte EMRFS-Committer und mehrteilige Uploads

Um den S3-optimierten EMRFS-Committer zu verwenden, müssen Sie mehrteilige Uploads in Amazon EMR aktivieren. Mehrteilige Uploads sind standardmäßig aktiviert. Sie können diese Option bei Bedarf erneut aktivieren. Weitere Informationen finden Sie unter [Konfigurieren von mehrteiligen Uploads für Amazon S3](#) im Verwaltungshandbuch für Amazon EMR.

Der S3-optimierte EMRFS-Committer verwendet die transaktionsähnlichen Merkmale von mehrteiligen Uploads, um sicherzustellen, dass Dateien, die beim Versuch, Aufgaben auszuführen

geschrieben werden, nur am Ausgabespeicherort des Auftrags angezeigt werden. Durch die Verwendung von mehrteiligen Uploads auf diese Weise verbessert der Committer die Leistung des Task-Commits im Vergleich zur Standardalgorithmusversion 2. `FileOutputCommitter` Wenn Sie den S3-optimierten EMRFS-Committer verwenden, gilt es einige wichtige Unterschiede zu dem herkömmlichen Verhalten bei mehrteiligen Uploads zu berücksichtigen:

- Mehrteilige Uploads werden immer ausgeführt, unabhängig von der Dateigröße. Dies unterscheidet sich vom Standardverhalten von EMRFS, bei dem die Eigenschaft `fs.s3n.multipart.uploads.split.size` die Dateigröße steuert, in der mehrteilige Uploads ausgelöst werden.
- Mehrteilige Uploads verbleiben für einen längeren Zeitraum in einem Status, in dem sie nicht abgeschlossen sind, bis die Aufgabe übertragen oder abgebrochen wird. Dies unterscheidet sich von der Standard-Verhalten von EMRFS. Dort wird ein mehrteiliger Upload abgeschlossen, wenn eine Aufgabe den Schreibvorgang für eine bestimmte Datei beendet hat.

Aufgrund dieser Unterschiede vergrößert sich bei mehrteiligen Uploads die Wahrscheinlichkeit, dass unvollständige mehrteilige Uploads zurückbleiben, wenn ein Spark Executor JVM abstürzt oder zerstört wird, während Aufgaben ausgeführt oder Daten auf Amazon S3 geschrieben werden. Aus diesem Grund sollten Sie bei Verwendung des S3-optimierten EMRFS-Committer darauf achten, den bewährten Methoden für die Verwaltung von fehlgeschlagenen mehrteiligen Uploads zu folgen. Weitere Informationen finden Sie unter [Bewährte Methoden](#) für die Arbeit mit Amazon-S3-Buckets im Verwaltungshandbuch für Amazon EMR.

## Überlegungen zur Auftragsoptimierung

Der S3-optimierte EMRFS-Committer verbraucht eine geringe Speichermenge für jede Datei, die versuchsweise von einer Aufgabe geschrieben wird, bis die Aufgabe übermittelt oder abgebrochen wird. Bei den meisten Aufträgen ist die Menge des belegten Speichers vernachlässigbar. Bei Aufträgen mit Aufgaben mit langer Ausführungsdauer, die eine große Anzahl von Dateien schreiben, macht sich der Arbeitsspeicher, den der Committer benötigt, bemerkbar, und dies erfordert möglicherweise Anpassungen an den für Spark Executor zugeteilten Arbeitsspeicher. Sie können die Eigenschaft `spark.executor.memory` verwenden, um den Executor-Arbeitsspeicher anzupassen. Als Faustregel gilt: für jeweils 100.000 Dateien, die eine einzelne Aufgabe schreiben, werden in der Regel zusätzlich 100 MB Arbeitsspeicher benötigt. Weiter Informationen finden Sie unter [Anwendungseigenschaften](#) in der Dokumentation zur Konfiguration von Apache Spark.

## Aktivieren Sie den S3-optimierten Committer für Amazon EMR 5.19.0

Wenn Sie Amazon EMR 5.19.0 verwenden, können Sie die Eigenschaft `spark.sql.parquet.fs.optimized.committer.optimization-enabled` manuell auf `true` einstellen, wenn Sie einen Cluster erstellen, oder aus Spark heraus, wenn Sie Amazon EMR verwenden.

Aktivieren des S3-optimierten EMRFS-Committers beim Erstellen eines Clusters

Verwenden Sie die Konfigurationsklassifizierung `spark-defaults`, um die `spark.sql.parquet.fs.optimized.committer.optimization-enabled`-Eigenschaften auf `true` festzulegen. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Aktivieren des S3-optimierten EMRFS-Committers über Spark

Sie können `spark.sql.parquet.fs.optimized.committer.optimization-enabled` auf `true` festlegen, indem Sie dies in einer `SparkConf` fest kodieren, als `--conf`-Parameter in der Spark-Shell übergeben, die Tools `spark-submit` bzw. `spark-sql` verwenden oder dies in `conf/spark-defaults.conf` angeben. Weiter Informationen finden Sie unter [Spark-Konfiguration](#) in der Apache-Spark-Dokumentation.

Im folgenden Beispiel wird gezeigt, wie Sie den Committer während der Ausführung eines `spark-sql`-Befehls aktivieren.

```
spark-sql \  
  --conf spark.sql.parquet.fs.optimized.committer.optimization-enabled=true \  
  -e "INSERT OVERWRITE TABLE target_table SELECT * FROM source_table;"
```

## Verwenden Sie das EMRFS-S3-optimierte Commit-Protokoll

Das EMRFS S3-optimierte Commit-Protokoll ist eine alternative [FileCommitProtocol](#) Implementierung, die für das Schreiben von Dateien mit dynamischem Spark-Partitionsüberschreiben in Amazon S3 optimiert ist, wenn EMRFS verwendet wird. Das Protokoll verbessert die Anwendungsleistung, indem Umbenennungsvorgänge in Amazon S3 während der Commit-Phase für das dynamische Überschreiben von Spark-Partitionen vermieden werden.

Beachten Sie, dass [EMRFS-S3-optimierte Committer verwenden](#) auch die Leistung verbessert, indem es Umbenennungsvorgänge vermeidet. Es funktioniert jedoch nicht für Fälle dynamischer Partitionsüberschreibungen, während die Verbesserungen des Commit-Protokolls nur auf Fälle dynamischer Partitionsüberschreibungen abzielen.

Die Committer-Klasse ist verfügbar für Amazon-EMR-Version 5.30.0 und höher und ist bei 6.2.0 und höher standardmäßig aktiviert. Amazon EMR hat ab Version 5.31.0 eine Verbesserung der Parallelität hinzugefügt. Das Protokoll wird für Spark-Jobs verwendet, die Spark SQL oder Datasets verwenden. DataFrames Es gibt Situationen, in denen das Commit Protokoll nicht verwendet wird. Weitere Informationen finden Sie unter [Anforderungen für das EMRFS-S3-optimierte Commit-Protokoll](#).

## Themen

- [Anforderungen für das EMRFS-S3-optimierte Commit-Protokoll](#)
- [Das S3-optimierte EMRFS-Commit-Protokoll und mehrteilige Uploads](#)
- [Überlegungen zur Auftragsoptimierung](#)


## Anforderungen für das EMRFS-S3-optimierte Commit-Protokoll

Das EMRFS-S3-optimierte Commit-Protokoll wird verwendet, wenn die folgenden Bedingungen erfüllt sind:

- Sie führen Spark-Jobs aus, die Spark SQL oder Datasets verwenden DataFrames, um partitionierte Tabellen zu überschreiben.
- Sie führen Spark-Jobs aus, deren Partitionsüberschreibmodus `dynamic` ist.
- In Amazon EMR sind mehrteilige Uploads aktiviert. Dies ist die Standardeinstellung. Weitere Informationen finden Sie unter [Das S3-optimierte EMRFS-Commit-Protokoll und mehrteilige Uploads](#).
- Der Dateisystem-Cache für EMRFS ist aktiviert. Dies ist die Standardeinstellung. Vergewissern Sie sich, dass die Einstellung `fs.s3.impl.disable.cache` auf `false` gesetzt ist.
- Die integrierte Datenquellenunterstützung von Spark wird verwendet. Die integrierte Datenquellenunterstützung wird in den folgenden Fällen verwendet:
  - Wenn Aufträge in integrierte Datenquellen oder Tabellen schreiben.
  - Wenn Aufträge in die Parquet-Tabelle des Hive-Metastores schreiben. Das passiert, wenn `spark.sql.hive.convertInsertingPartitionedTable` und `spark.sql.hive.convertMetastoreParquet` beide auf „wahr“ gesetzt sind. Dies sind die Standardeinstellungen.
  - Wenn Aufträge in die ORC-Tabelle des Hive-Metastores schreiben. Das passiert, wenn `spark.sql.hive.convertInsertingPartitionedTable` und

`spark.sql.hive.convertMetastoreOrc` beide auf `true` gesetzt sind. Dies sind die Standardeinstellungen.

- Spark-Auftrags-Vorgänge, die z. B. in einen Standard-Partitionsspeicherort schreiben, z. B. `${table_location}/k1=v1/k2=v2/`, verwenden das Commit-Protokoll. Das Protokoll wird nicht verwendet, wenn ein Jobvorgang in einen benutzerdefinierten Partitionsspeicherort schreibt, beispielsweise wenn ein benutzerdefinierter Partitionsspeicherort mit dem Befehl `ALTER TABLE SQL` festgelegt wird.
- Für Spark müssen die folgenden Werte verwendet werden:
  - muss `spark.sql.sources.commitProtocolClass` auf `org.apache.spark.sql.execution.datasources.SQLEmrOptimizedCommitProtocol` festgelegt sein. Dies ist die Standardeinstellung für Amazon-EMR-Versionen 5.30.0 und höher sowie 6.2.0 und höher.
  - Die Schreiboption `partitionOverwriteMode` oder `spark.sql.sources.partitionOverwriteMode` muss auf `dynamic` gesetzt sein. Die Standardeinstellung lautet `static`.

 Note

Die Schreiboption `partitionOverwriteMode` wurde mit Spark 2.4.0 eingeführt. Legen Sie für Spark-Version 2.3.2, die in Amazon-EMR-Version 5.19.0 enthalten ist, die Eigenschaft `spark.sql.sources.partitionOverwriteMode` fest.

- Wenn Spark-Aufträge in die Hive-Metastore-Parquet-Tabelle überschreiben muss `spark.sql.hive.convertMetastoreParquet`, `spark.sql.hive.convertInsertingPartitionedTable` und `spark.sql.hive.convertMetastore.partitionOverwriteMode` auf `true` gesetzt werden. Dies sind die Standardeinstellungen.
- Wenn Spark-Aufträge in die Hive-Metastore-ORC-Tabelle überschreiben muss `spark.sql.hive.convertMetastoreOrc`, `spark.sql.hive.convertInsertingPartitionedTable` und `spark.sql.hive.convertMetastore.partitionOverwriteMode` auf `true` gesetzt werden. Dies sind die Standardeinstellungen.

## Example – Dynamischer Partitionsüberschreibmodus

In diesem Scala-Beispiel wird die Optimierung ausgelöst. Als erstes setzen Sie die `partitionOverwriteMode`-Eigenschaft auf `dynamic`. Dadurch werden nur die Partitionen überschrieben, auf die Sie Daten schreiben. Anschließend geben Sie dynamische Partitionsspalten mit `partitionBy` an und legen den Schreibmodus auf `overwrite` fest.

```
val dataset = spark.range(0, 10)
  .withColumn("dt", expr("date_sub(current_date(), id)"))

dataset.write.mode("overwrite")           // "overwrite" instead of "insert"
  .option("partitionOverwriteMode", "dynamic") // "dynamic" instead of "static"
  .partitionBy("dt")                       // partitioned data instead of
  unpartitioned data
  .parquet("s3://EXAMPLE-D0C-BUCKET/output") // "s3://" to use Amazon EMR file
  system, instead of "s3a://" or "hdfs://"
```

Wenn das EMRFS-S3-optimierte Commit-Protokoll nicht verwendet wird

Im Allgemeinen funktioniert das für EMRFS S3 optimierte Commit-Protokoll

genauso wie das Open-Source-Standard-Spark-SQL-Commit-Protokoll

`org.apache.spark.sql.execution.datasources.SQLHadoopMapReduceCommitProtocol`.

In den folgenden Situationen wird keine Optimierung durchgeführt.

| Situation                                                            | Warum das Commit-Protokoll nicht verwendet wird                                                |
|----------------------------------------------------------------------|------------------------------------------------------------------------------------------------|
| Wenn Sie in HDFS schreiben                                           | Das Commit-Protokoll unterstützt nur das Schreiben in Amazon S3 mit EMRFS-Technologie.         |
| Bei Verwendung des S3A-Dateisystems                                  | Das Commit-Protokoll unterstützt nur EMRFS.                                                    |
| Wenn Sie die RDD-API von Spark verwenden<br>MapReduce                | Das Commit-Protokoll unterstützt nur die Verwendung von SparkSQL- DataFrame oder Dataset-APIs. |
| Wenn das dynamische Überschreiben der Partition nicht ausgelöst wird | Das Commit-Protokoll optimiert nur Fälle, in denen dynamische Partitionen überschrieben        |



| Situation | Warum das Commit-Protokoll nicht verwendet wird                                          |
|-----------|------------------------------------------------------------------------------------------|
|           | werden. Für andere Fälle siehe <a href="#">EMRFS-S3-optimierte Committer verwenden</a> . |

Die folgenden Scala-Beispiele veranschaulichen einige zusätzliche Situationen, die das EMRFS-S3-optimierte Commit-Protokoll an `SQLHadoopMapReduceCommitProtocol` delegiert.

#### Example – Dynamischer Partitionsüberschreibmodus mit benutzerdefiniertem Partitionsspeicherort

In diesem Beispiel überschreiben die Scala-Programme zwei Partitionen im dynamischen Partitionsüberschreibmodus. Eine Partition verfügt über einen benutzerdefinierten Partitionsspeicherort. Die andere Partition verwendet den Standard-Partitionsspeicherort. Das S3-optimierte EMRFS-Commit-Protokoll wird nur zum Verbessern der Aufgabenausgabe in die Partition genutzt, die den Standard-Partitionsspeicherort verwendet.

```
val table = "dataset"
val inputView = "tempView"
val location = "s3://bucket/table"

spark.sql(s"""
  CREATE TABLE $table (id bigint, dt date)
  USING PARQUET PARTITIONED BY (dt)
  LOCATION '$location'
""")

// Add a partition using a custom location
val customPartitionLocation = "s3://bucket/custom"
spark.sql(s"""
  ALTER TABLE $table ADD PARTITION (dt='2019-01-28')
  LOCATION '$customPartitionLocation'
""")

// Add another partition using default location
spark.sql(s"ALTER TABLE $table ADD PARTITION (dt='2019-01-29')")

def asDate(text: String) = lit(text).cast("date")

spark.range(0, 10)
  .withColumn("dt",
```

```
when($"id" > 4, asDate("2019-01-28")).otherwise(asDate("2019-01-29")))
.createTempView(inputView)

// Set partition overwrite mode to 'dynamic'
spark.sql(s"SET spark.sql.sources.partitionOverwriteMode=dynamic")

spark.sql(s"INSERT OVERWRITE TABLE $table SELECT * FROM $inputView")
```

Der Scala-Code erstellt die folgenden Amazon-S3-Objekte:

```
custom/part-00001-035a2a9c-4a09-4917-8819-e77134342402.c000.snappy.parquet
custom_${folder$}
table/_SUCCESS
table/dt=2019-01-29/part-00000-035a2a9c-4a09-4917-8819-e77134342402.c000.snappy.parquet
table/dt=2019-01-29_${folder$}
table_${folder$}
```

#### Note

Das Schreiben in benutzerdefinierte Partitionsspeicherorte in früheren Spark-Versionen kann zu Datenverlust führen. In diesem Beispiel `dt= '2019-01-28'` würde die Partition verloren gehen. Weitere Informationen finden Sie unter [SPARK-35106](#). Dieses Problem wurde in Amazon-EMR-Version 5.33.0 und höher behoben, ausgenommen 6.0.x und 6.1.x.

Beim Schreiben in Partitionen an benutzerdefinierten Speicherorten verwendet Spark einen Commit-Algorithmus, ähnlich wie im vorherigen Beispiel. Dies wird im Folgenden beschrieben. Genau wie bei dem vorherigen Beispiel führt der Algorithmus zu sequenziellen Umbenennungen, wodurch die Leistung beeinträchtigt werden kann.

Der Algorithmus in Spark 2.4.0 folgt diesen Schritten:

1. Beim Schreiben von Ausgabe in eine Partition an einem benutzerdefinierten Speicherort werden Aufgaben in eine Datei unter dem Staging-Verzeichnis von Spark geschrieben, das unter dem endgültigen Ausgabespeicherort erstellt wird. Der Name der Datei enthält eine zufällige UUID zum Schutz vor Datei-Kollisionen. Der Aufgabe-Versuch verfolgt jede Datei zusammen mit dem gewünschten endgültigen Ausgabepfad nach.
2. Wenn eine Aufgabe erfolgreich abgeschlossen wird, stellt sie dem Treiber die Dateien und die für sie gewünschten endgültigen Ausgabepfade bereit.

3. Nachdem alle Aufgaben beendet wurden, benennt die Auftrags-Commit-Phase alle Dateien, die für Partitionen an benutzerdefinierten Speicherorten geschrieben wurden, sequentiell in ihre endgültigen Ausgabepfade um.
4. Das Staging-Verzeichnis wird gelöscht, bevor der Auftrags-Commit-Phase abgeschlossen ist.

## Das S3-optimierte EMRFS-Commit-Protokoll und mehrteilige Uploads

Um die Optimierung für dynamisches Partitionsüberschreiben im EMRFS-S3-optimierten Commit-Protokoll nutzen zu können, müssen mehrteilige Uploads in Amazon EMR aktiviert sein. Mehrteilige Uploads sind standardmäßig aktiviert. Sie können diese Option bei Bedarf erneut aktivieren. Weitere Informationen finden Sie unter [Konfigurieren von mehrteiligen Uploads für Amazon S3](#) im Verwaltungshandbuch für Amazon EMR.

Beim dynamischen Überschreiben der Partition nutzt das für EMRFS S3 optimierte Commit-Protokoll die transaktionsähnlichen Eigenschaften mehrteiliger Uploads, um sicherzustellen, dass Dateien, die durch Aufgabenversuche geschrieben wurden, beim Auftrags-Commit nur am Ausgabeort des Auftrags angezeigt werden. Durch die Verwendung von mehrteiligen Uploads auf diese Weise verbessert das Commit-Protokoll die Leistung von Auftrags-Commits im Vergleich zum Standard-`SQLHadoopMapReduceCommitProtocol`. Wenn Sie das S3-optimierte EMRFS-Commit-Protokoll verwenden, gilt es einige wichtige Unterschiede zu dem herkömmlichen Verhalten bei mehrteiligen Uploads zu berücksichtigen:

- Mehrteilige Uploads werden immer ausgeführt, unabhängig von der Dateigröße. Dies unterscheidet sich vom Standardverhalten von EMRFS, bei dem die Eigenschaft `fs.s3n.multipart.uploads.split.size` die Dateigröße steuert, in der mehrteilige Uploads ausgelöst werden.
- Mehrteilige Uploads verbleiben für einen längeren Zeitraum in einem Status, in dem sie nicht abgeschlossen sind, bis die Aufgabe übertragen oder abgebrochen wird. Dies unterscheidet sich von der Standard-Verhalten von EMRFS. Dort wird ein mehrteiliger Upload abgeschlossen, wenn eine Aufgabe den Schreibvorgang für eine bestimmte Datei beendet hat.

Aufgrund dieser Unterschiede vergrößert sich bei mehrteiligen Uploads die Wahrscheinlichkeit, dass unvollständige mehrteilige Uploads zurückbleiben, wenn ein Spark Executor JVM abstürzt oder zerstört wird, während Aufgaben ausgeführt oder Daten auf Amazon S3 oder Spark Driver JVM geschrieben werden. Aus diesem Grund sollten Sie bei Verwendung des S3-optimierten EMRFS-Committer darauf achten, den bewährten Methoden für die Verwaltung von fehlgeschlagenen

mehrteiligen Uploads zu folgen. Weitere Informationen finden Sie unter [Bewährte Methoden](#) für die Arbeit mit Amazon-S3-Buckets im Verwaltungshandbuch für Amazon EMR.

## Überlegungen zur Auftragsoptimierung

Auf Spark-Ausführern verbraucht das für EMRFS-S3-optimierte Commit-Protokoll eine geringe Speichermenge für jede Datei, die versuchsweise von einer Aufgabe geschrieben wird, bis die Aufgabe übermittelt oder abgebrochen wird. Bei den meisten Aufträgen ist die Menge des belegten Speichers vernachlässigbar.

Auf Spark-Treibern benötigt das für EMRFS-S3-optimierte Commit-Protokoll Speicher, um die Metadateninformationen jeder übergebenen Datei zu speichern, bis der Auftrag festgeschrieben oder abgebrochen wird. Bei den meisten Aufträgen ist die standardmäßige Speichereinstellung des Spark-Treibers vernachlässigbar.

Bei Aufträgen mit Aufgaben mit langer Laufzeit, die eine große Anzahl von Dateien schreiben, kann der Speicherverbrauch des Commit-Protokolls spürbar sein und Anpassungen des für Spark zugewiesenen Speichers erfordern, insbesondere für Spark-Ausführer. Sie können den Speicher mithilfe der `spark.driver.memory`-Eigenschaft für Spark-Treiber und der Eigenschaft für `spark.executor.memory`-Spark-Aufträge optimieren. Als Faustregel gilt: für jeweils 100.000 Dateien, die eine einzelne Aufgabe schreiben, werden in der Regel zusätzlich 100 MB Arbeitsspeicher benötigt. Weiter Informationen finden Sie unter [Anwendungseigenschaften](#) in der Dokumentation zur Konfiguration von Apache Spark.

## Amazon-S3-Anforderungen mit EMRFS wiederholen

Dieses Thema enthält Informationen zu den Wiederholungsstrategien, die Sie verwenden können, wenn Sie Anforderungen an Amazon S3 mit EMRFS stellen. Wenn Ihre Anforderungsrate steigt, versucht S3, zu skalieren, um die neue Rate zu unterstützen. Während dieses Vorgangs kann S3 Anforderungen drosseln und einen `503 Slow Down`-Fehler zurückgeben. Um die Erfolgsquote Ihrer S3-Anforderungen zu verbessern, können Sie Ihre Wiederholungsstrategie anpassen, indem Sie Eigenschaften in Ihrer `emrfs-site`-Konfiguration konfigurieren.

Sie können Ihre Wiederholungsstrategie wie folgt anpassen.

- Erhöhen Sie das maximale Wiederholungslimit für die standardmäßige exponentielle Backoff-Wiederholungsstrategie.
- Aktivieren und konfigurieren Sie die AIMD-Wiederholungsstrategie (additive-increase/multiplicative-decrease). AIMD wird für Amazon-EMR-Versionen 6.4.0 und höher unterstützt.

## Verwenden Sie die standardmäßige exponentielle Backoff-Strategie

Standardmäßig verwendet EMRFS eine exponentielle Backoff-Strategie, um Amazon-S3-Anfragen erneut zu versuchen. Das standardmäßige EMRFS-Wiederholungslimit ist 15. Um einen 503 Slow Down-S3-Fehler zu vermeiden, können Sie das Wiederholungslimit erhöhen, wenn Sie einen neuen Cluster, in einem laufenden Cluster oder zur Laufzeit der Anwendung erstellen.

Um das Wiederholungslimit zu erhöhen, müssen Sie den Wert für `fs.s3.maxRetries` in Ihrer `emrfs-site`-Konfiguration ändern. Die folgende Beispielkonfiguration setzt `fs.s3.maxRetries` auf einen benutzerdefinierten Wert von 30.

```
[
  {
    "Classification": "emrfs-site",
    "Properties": {
      "fs.s3.maxRetries": "30"
    }
  }
]
```

Weitere Informationen zum Arbeiten mit Konfigurationsobjekten finden Sie unter [Anwendungen konfigurieren](#).

## Verwenden Sie die AIMD-Wiederholungsstrategie

Mit Amazon-EMR-Version 6.4.0 und höher unterstützt EMRFS eine alternative Wiederholungsstrategie, die auf einem AIMD-Modell (Additive-Increase/Multiplicative-Decrease) basiert. Die AIMD-Wiederholungsstrategie ist besonders nützlich, wenn Sie mit großen Amazon-EMR-Clustern arbeiten.

AIMD berechnet anhand von Daten über kürzlich erfolgreiche Anfragen eine benutzerdefinierte Anforderungsrate. Diese Strategie verringert die Anzahl der gedrosselten Anforderungen und die Gesamtzahl der pro Anforderungen erforderlichen Versuche.

Um die AIMD-Wiederholungsstrategie zu aktivieren, müssen Sie die `fs.s3.aimd.enabled`-Eigenschaft in `true` in Ihrer `emrfs-site`-Konfiguration einstellen, wie im folgenden Beispiel gezeigt.

```
[
  {
```

```

    "Classification": "emrfs-site",
    "Properties": {
      "fs.s3.aimd.enabled": "true"
    }
  }
]

```

Weitere Informationen zum Arbeiten mit Konfigurationsobjekten finden Sie unter [Anwendungen konfigurieren](#).

## Erweiterte Einstellungen für AIMD-Wiederholungen

Sie können die in der folgenden Tabelle aufgeführten Eigenschaften konfigurieren, um das Wiederholungsverhalten zu verfeinern, wenn Sie die AIMD-Wiederholungsstrategie verwenden. Für die meisten Anwendungsfälle empfehlen wir die Verwendung der Standardeinstellung.

### Erweiterte Eigenschaften der AIMD-Wiederholungsstrategie

| Eigenschaft                               | Standardwert | Beschreibung                                                                                                                                         |
|-------------------------------------------|--------------|------------------------------------------------------------------------------------------------------------------------------------------------------|
| <code>fs.s3.aimd.increaseIncrement</code> | 0.1          | Steuert, wie schnell die Anforderungsrate steigt, wenn aufeinanderfolgende Anforderungen erfolgreich sind.                                           |
| <code>fs.s3.aimd.reductionFactor</code>   | 2            | Steuert, wie schnell die Anforderungsrate sinkt, wenn Amazon S3 eine 503-Antwort zurückgibt. Der Standardfaktor von 2 halbiert die Anforderungsrate. |
| <code>fs.s3.aimd.minRate</code>           | 0.1          | Legt die Untergrenze für die Anforderungsrate fest, wenn Anfragen dauerhaft durch S3 gedrosselt werden.                                              |
| <code>fs.s3.aimd.initialRate</code>       | 5500         | Legt die anfängliche Anforderungsrate fest, die sich dann                                                                                            |

| Eigenschaft                          | Standardwert | Beschreibung                                                                                                                                                                                                                                              |
|--------------------------------------|--------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                                      |              | entsprechend den Werten <code>fs.s3.aimd.increaseIncrement</code> und <code>fs.s3.aimd.reductionFactor</code> angeben.<br><br>Die anfängliche Rate wird auch für GET-Anforderungen verwendet und für PUT-Anforderungen proportional (3500/5500) skaliert. |
| <code>fs.s3.aimd.adjustWindow</code> | 2            | Steuert, wie oft die Anforderungsrate angepasst wird, gemessen an der Anzahl der Antworten.                                                                                                                                                               |
| <code>fs.s3.aimd.maxAttempts</code>  | 100          | Legt die maximale Anzahl der Versuche fest, eine Anforderung zu versuchen.                                                                                                                                                                                |

## Einen Spark-Schritt hinzufügen

Sie können mit Amazon-EMR-Schritten Arbeit an das auf einem EMR-Cluster installierte Spark-Framework senden. Weitere Informationen finden Sie unter [Schritte](#) im Verwaltungshandbuch für Amazon EMR. In der Konsole und CLI erfolgt dies mithilfe eines Spark-Anwendungsschritts, der das Skript `spark-submit` als Schritt in Ihrem Namen ausführt. In der API rufen Sie `spark-submit` mithilfe eines Schritts auf und verwenden `command-runner.jar`.

Weitere Informationen zum Senden von Anwendungen an Spark finden Sie im Thema [Bewerbungen einreichen](#) in der Apache-Spark-Dokumentation.

So senden Sie einen Spark-Schritt mithilfe der Konsole

1. Öffnen Sie die Amazon EMR-Konsole unter <https://console.aws.amazon.com/emr>.
2. Wählen Sie unter Cluster List den Namen Ihres Clusters aus.

3. Scrollen Sie zum Abschnitt Steps (Schritte) und erweitern Sie ihn. Wählen Sie anschließend Add step (Schritt hinzufügen) aus.
4. Gehen Sie im Dialogfeld Add Step folgendermaßen vor:
  - Wählen Sie für Step type die Option Spark application aus.
  - Übernehmen Sie unter Name den Standardnamen ("Spark application") oder geben Sie einen neuen Namen ein.
  - Wählen Sie unter Deploy mode (Bereitstellungsmodus) den Modus Client oder Cluster aus. Der Client-Modus startet das Treiberprogramm in der Primär-Instance des Clusters, während der Cluster-Modus das Treiberprogramm im Cluster startet. Im Clientmodus wird die Protokollausgabe des Treibers in den Schrittprotokollen angezeigt, während im Clustermodus die Protokollausgabe des Treibers in den Protokollen für den ersten YARN-Container angezeigt wird. Weitere Informationen finden Sie unter [Übersicht über den Clustermodus](#) in der Apache-Spark-Dokumentation.
  - Geben Sie die gewünschten spark-submit-Optionen an. Weitere Informationen zu den spark-submit-Optionen finden Sie unter [Starten von Anwendungen mit Spark-Submit](#).
  - Geben Sie bei Application location (Anwendungsspeicherort) den lokalen oder S3-URI-Pfad der Anwendung an.
  - Lassen Sie unter Arguments (Parameter) das Feld leer.
  - Übernehmen Sie unter Action on failure (Aktion bei Fehler) die Standardeinstellung Continue (Fortfahren).
5. Wählen Sie Hinzufügen aus. Der Schritt wird in der Konsole mit dem Status "Pending" angezeigt.
6. Bei Ausführung des Schritts ändert sich dessen Status von Pending zu Running und Completed. Wählen Sie über der Spalte Actions (Aktionen) das Symbol Refresh (Ansicht aktualisieren) aus, um den Status zu aktualisieren.
7. Die Ergebnisse des Schritts stehen in der Amazon-EMR-Konsole auf der Seite Cluster Details neben dem Schritt unter Protokolldateien zur Verfügung, wenn Sie die Protokollierung konfiguriert haben. Optional finden Sie Informationen zum Schritt im Protokoll-Bucket, den Sie beim Start des Clusters konfiguriert haben.

Um Arbeiten bei Spark einzureichen, verwenden Sie den AWS CLI

Senden Sie einen Schritt, wenn Sie den Cluster erstellen, oder verwenden Sie den Unterbefehl `aws emr add-steps` in einem vorhandenen Cluster.



## 1. Verwenden Sie `create-cluster`, wie im folgenden Beispiel gezeigt.

### Note

Linux-Zeilenfortsetzungszeichen (`\`) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (`^`).

```
aws emr create-cluster --name "Add Spark Step Cluster" --release-label emr-7.1.0 --
applications Name=Spark \
--ec2-attributes KeyName=myKey --instance-type m5.xlarge --instance-count 3 \
--steps Type=Spark,Name="Spark Program",ActionOnFailure=CONTINUE,Args=[--
class,org.apache.spark.examples.SparkPi,/usr/lib/spark/examples/jars/spark-
examples.jar,10] --use-default-roles
```

Alternativ können Sie auch `command-runner.jar` verwenden, wie im folgenden Beispiel gezeigt.

```
aws emr create-cluster --name "Add Spark Step Cluster" --release-label emr-7.1.0 \
--applications Name=Spark --ec2-attributes KeyName=myKey --instance-type m5.xlarge
--instance-count 3 \
--steps Type=CUSTOM_JAR,Name="Spark Program",Jar="command-
runner.jar",ActionOnFailure=CONTINUE,Args=[spark-example,SparkPi,10] --use-default-
roles
```

### Note

Linux-Zeilenfortsetzungszeichen (`\`) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (`^`).

## 2. Alternativ können Sie einem bereits ausgeführten Cluster Schritte hinzufügen. Verwenden Sie `add-steps`.

```
aws emr add-steps --cluster-id j-2AXXXXXXGAPLF --steps
Type=Spark,Name="Spark Program",ActionOnFailure=CONTINUE,Args=[--
```

```
class,org.apache.spark.examples.SparkPi,/usr/lib/spark/examples/jars/spark-  
examples.jar,10]
```

Alternativ können Sie auch `command-runner.jar` verwenden, wie im folgenden Beispiel gezeigt.

```
aws emr add-steps --cluster-id j-2AXXXXXXGAPLF --steps Type=CUSTOM_JAR,Name="Spark  
Program",Jar="command-runner.jar",ActionOnFailure=CONTINUE,Args=[spark-  
example,SparkPi,10]
```

So senden Sie Arbeit mithilfe des SDK für Java an Spark

1. Das folgende Beispiel zeigt, wie Sie einen Schritt zu einem Cluster mit Spark mittels Java hinzufügen.

```
AWSCredentials credentials = new BasicAWSCredentials(accessKey, secretKey);  
AmazonElasticMapReduce emr = new AmazonElasticMapReduceClient(credentials);  
  
StepFactory stepFactory = new StepFactory();  
AmazonElasticMapReduceClient emr = new AmazonElasticMapReduceClient(credentials);  
AddJobFlowStepsRequest req = new AddJobFlowStepsRequest();  
req.withJobFlowId("j-1K48XXXXXXHCB");  
  
List<StepConfig> stepConfigs = new ArrayList<StepConfig>();  
  
HadoopJarStepConfig sparkStepConf = new HadoopJarStepConfig()  
    .withJar("command-runner.jar")  
    .withArgs("spark-submit", "--executor-memory", "1g", "--  
class", "org.apache.spark.examples.SparkPi", "/usr/lib/spark/examples/jars/spark-  
examples.jar", "10");  
  
StepConfig sparkStep = new StepConfig()  
    .withName("Spark Step")  
    .withActionOnFailure("CONTINUE")  
    .withHadoopJarStep(sparkStepConf);  
  
stepConfigs.add(sparkStep);  
req.withSteps(stepConfigs);  
AddJobFlowStepsResult result = emr.addJobFlowSteps(req);
```

2. Untersuchen Sie die Protokolle für den Schritt, um die Ergebnisse einzusehen. Sie können dies in der tun, AWS Management Console sofern Sie die Protokollierung aktiviert haben, indem Sie Schritte wählen, Ihren Schritt auswählen und dann für Protokolldateien entweder `stdout` oder wählen `stderr`. Wählen Sie View Logs aus, um eine Ansicht der verfügbaren Protokoll zu erhalten.

## Überschreiben der standardmäßigen Spark-Konfigurationseinstellungen

Vielleicht möchten Sie die standardmäßigen Spark-Konfigurationswerte je nach Anwendung überschreiben. Sie können dies tun, wenn Sie Anwendungen mithilfe eines Schritts senden, wodurch im Wesentlichen Optionen an `spark-submit` übergeben werden. Sie möchten z. B. den Speicher ändern, der einem Executor-Prozess zugeordnet wurde. Dazu modifizieren Sie `spark.executor.memory`. Dann geben Sie den Schalter `--executor-memory` mit einem Argument wie dem folgenden an:

```
spark-submit --executor-memory 1g --class org.apache.spark.examples.SparkPi /usr/lib/spark/examples/jars/spark-examples.jar 10
```

Auf ähnliche Weise können Sie `--executor-cores` und `--driver-memory` optimieren. In einem Schritt würden Sie die folgenden Argumente bereitstellen:

```
--executor-memory 1g --class org.apache.spark.examples.SparkPi /usr/lib/spark/examples/jars/spark-examples.jar 10
```

Sie können mithilfe der Option `--conf` auch Einstellungen optimieren, die eventuell über keinen integrierten Schalter verfügen. Weitere Informationen zu anderen optimierbaren Einstellungen finden Sie im Thema [Dynamisches Laden von Spark-Eigenschaften](#) in der Apache-Spark-Dokumentation.

## Anzeigen des Spark-Anwendungsverlaufs

Sie können Spark-, YARN-Anwendungs- und Tez-UI-Details über die Registerkarte Anwendungsbenutzeroberflächen auf der Detailseite eines Clusters in der Konsole anzeigen. Die Benutzeroberflächen (UI) der Amazon-EMR-Anwendung erleichtern Ihnen die Fehlerbehebung und Analyse aktiver Jobs und des Jobverlaufs.

Weitere Informationen finden Sie unter [Anwendungsverlauf anzeigen](#) im Verwaltungshandbuch für Amazon EMR.

## Zugreifen auf Web-Benutzerschnittstellen in Spark

Sie können die Spark-Web-Benutzeroberflächen anzeigen, indem Sie die Verfahren zum Erstellen eines SSH-Tunnels oder zum Erstellen eines Proxys im Abschnitt [Connect to the Cluster](#) im Amazon EMR Management Guide befolgen und dann zum YARN ResourceManager für Ihren Cluster navigieren. Wählen Sie den Link unter Tracking UI für Ihre Anwendung aus. Wenn Ihre Anwendung läuft, sehen Sie. ApplicationMaster Auf diese Weise gelangen Sie zur Web-Benutzerschnittstelle der Spark-Anwendung auf Port 20888, wo auch immer sich der Treiber befindet. Der Treiber kann sich auf dem Primärknoten des Clusters befinden, wenn Sie den YARN-Clientmodus ausführen. Wenn Sie eine Anwendung im YARN-Clustermodus ausführen, befindet sich der Treiber im ApplicationMaster Ordner für die Anwendung auf dem Cluster. Wenn Ihre Anwendung abgeschlossen ist, wird Verlauf angezeigt, der Sie zur HistoryServer Spark-UI-Portnummer 18080 des primären Knotens des EMR-Clusters weiterleitet. Dies gilt für Anwendungen, die bereits abgeschlossen sind. Sie können auch direkt zur HistoryServer Spark-Benutzeroberfläche unter `http://:18080/` navigieren. *master-public-dns-name*

Mit Amazon-EMR-Version 5.25.0 und höher können Sie über die Konsole auf die Benutzeroberfläche des Spark-Historie-Servers zugreifen, ohne einen Web-Proxy über eine SSH-Verbindung einzurichten. Weitere Informationen finden Sie unter [Anzeigen von persistenten Anwendungsbenutzeroberflächen](#).

## Verwenden des Amazon Kinesis Data Streams-Connectors für strukturiertes Streaming mit Spark

Die Amazon EMR-Versionen 7.1.0 und höher enthalten im Release-Image einen strukturierten Spark-Streaming-Konnektor für Amazon Kinesis Data Streams. Mit diesem Connector können Sie Spark auf Amazon EMR verwenden, um Daten zu verarbeiten, die in Amazon Kinesis Data Streams gespeichert sind. Der Connector unterstützt sowohl Verbrauchertypen GetRecords (gemeinsamer Durchsatz) als auch SubscribeToShard (erweiterter Fan-Out). Diese Integration basiert auf dem [spark-sql-kinesis-connector](#). Einzelheiten zu den ersten Schritten mit der Verwendung des Connectors finden Sie in der [README-Datei](#).

Das folgende Beispiel zeigt, wie Sie den Connector verwenden, um eine Spark-Anwendung mit Amazon EMR zu starten.

```
spark-submit my_kinesis_streaming_script.py
```

# Amazon-Redshift-Integration für Apache Spark mit Amazon EMR verwenden

Mit Amazon-EMR-Version 6.4.0 und höher enthält jedes Versions-Image einen Konnektor zwischen [Apache Spark](#) und Amazon Redshift. Mit diesem Konnektor können Sie Spark auf Amazon EMR verwenden, um in Amazon Redshift gespeicherte Daten zu verarbeiten. Für die Amazon-EMR-Versionen 6.4.0 bis 6.8.0 basiert die Integration auf dem [spark-redshift-Open-Source-Konnektor](#). Für Amazon-EMR-Versionen 6.9.0 und höher wurde die [Amazon-Redshift-Integration für Apache Spark](#) von der Community-Version auf eine native Integration migriert.

## Themen

- [Starten einer Spark-Anwendung mithilfe der Amazon-Redshift-Integration für Apache Spark](#)
- [Mit der Amazon-Redshift-Integration für Apache Spark authentifizieren](#)
- [Lesen und Schreiben von und zu Amazon Redshift](#)
- [Überlegungen und Einschränkungen bei der Verwendung des Spark-Connectors](#)

## Starten einer Spark-Anwendung mithilfe der Amazon-Redshift-Integration für Apache Spark

Für die Amazon-EMR-Versionen 6.4 bis 6.9 müssen Sie die Option `--jars` oder `--packages` verwenden, um anzugeben, welche der folgenden JAR-Dateien Sie verwenden möchten. Die `--jars`-Option gibt Abhängigkeiten an, die lokal, in HDFS oder unter Verwendung von HTTP/S gespeichert sind. Weitere von der `--jars`-Option unterstützte Dateispeicherorte finden Sie unter [Erweitertes Abhängigkeitsmanagement](#) in der Spark-Dokumentation. Die `--packages`-Option spezifiziert Abhängigkeiten, die im öffentlichen Maven-Repository gespeichert sind.

- `spark-redshift.jar`
- `spark-avro.jar`
- `RedshiftJDBC.jar`
- `minimal-json.jar`

Amazon-EMR-Versionen 6.10.0 und höher erfordern die `minimal-json.jar`-Abhängigkeit nicht und installieren die anderen Abhängigkeiten standardmäßig automatisch in jedem Cluster. Die

folgenden Beispiele zeigen, wie Sie eine Spark-Anwendung mit der Amazon-Redshift-Integration für Apache Spark starten.

#### Amazon EMR 6.10.0 +

Das folgende Beispiel zeigt, wie eine Spark-Anwendung mit dem `spark-redshift`-Konnektor mit Amazon-EMR-Versionen 6.10 und höher gestartet wird.

```
spark-submit my_script.py
```

#### Amazon EMR 6.4.0 - 6.9.x

Um eine Spark-Anwendung mit dem `spark-redshift`-Konnektor auf den Amazon-EMR-Versionen 6.4 bis 6.9 zu starten, müssen Sie die Option `--jars` oder `--packages` verwenden, wie das folgende Beispiel zeigt. Beachten Sie, dass die mit der `--jars`-Option aufgeführten Pfade die Standardpfade für die JAR-Dateien sind.

```
spark-submit \  
  --jars /usr/share/aws/redshift/jdbc/RedshiftJDBC.jar,/usr/share/aws/redshift/  
  spark-redshift/lib/spark-redshift.jar,/usr/share/aws/redshift/spark-redshift/lib/  
  spark-avro.jar,/usr/share/aws/redshift/spark-redshift/lib/minimal-json.jar \  
  my_script.py
```

## Mit der Amazon-Redshift-Integration für Apache Spark authentifizieren

Wird verwendet AWS Secrets Manager , um Anmeldeinformationen abzurufen und eine Verbindung zu Amazon Redshift herzustellen

Das folgende Codebeispiel zeigt, wie Sie Anmeldeinformationen abrufen können AWS Secrets Manager , um eine Verbindung zu einem Amazon Redshift Redshift-Cluster mit der PySpark Schnittstelle für Apache Spark in Python herzustellen.

```
from pyspark.sql import SQLContext  
import boto3  
  
sc = # existing SparkContext  
sql_context = SQLContext(sc)  
  
secretsmanager_client = boto3.client('secretsmanager')
```

```
secret_manager_response = secretsmanager_client.get_secret_value(
    SecretId='string',
    VersionId='string',
    VersionStage='string'
)
username = # get username from secret_manager_response
password = # get password from secret_manager_response
url = "jdbc:redshift://redshifthost:5439/database?user=" + username + "&password=" +
    password

# Read data from a table
df = sql_context.read \
    .format("io.github.spark_redshift_community.spark.redshift") \
    .option("url", url) \
    .option("dbtable", "my_table") \
    .option("tempdir", "s3://path/for/temp/data") \
    .load()
```

## Verwenden Sie IAM, um Anmeldeinformationen abzurufen und eine Verbindung zu Amazon Redshift herzustellen

Sie können den von Amazon Redshift bereitgestellten JDBC-Version 2 verwenden, um mit dem Spark-Konnektor eine Verbindung mit Amazon Redshift herzustellen. Um AWS Identity and Access Management (IAM) zu verwenden, [konfigurieren Sie Ihre JDBC-URL](#) für die Verwendung der IAM-Authentifizierung. Um eine Verbindung zu einem Redshift-Cluster von Amazon EMR herzustellen, müssen Sie Ihrer IAM-Rolle die Berechtigung erteilen, temporäre IAM-Anmeldeinformationen abzurufen. Weisen Sie Ihrer IAM-Rolle die folgenden Berechtigungen zu, damit sie Anmeldeinformationen abrufen und Amazon-S3-Vorgänge ausführen kann.

- [Redshift: GetClusterCredentials](#) (für bereitgestellte Amazon Redshift Redshift-Cluster)
- [Redshift: DescribeClusters](#) (für bereitgestellte Amazon Redshift Redshift-Cluster)
- [Redshift: GetWorkgroup](#) (für serverlose Amazon Redshift Redshift-Arbeitsgruppen)
- [Redshift: GetCredentials](#) (für serverlose Amazon Redshift Redshift-Arbeitsgruppen)
- [s3: GetBucket](#)
- [s3: GetBucketLocation](#)
- [s3: GetObject](#)
- [s3: PutObject](#)
- [s3: GetBucketLifecycleConfiguration](#)

Weitere Informationen zu `GetClusterCredentials` finden Sie unter [Ressourcenrichtlinien für `GetClusterCredentials`](#).

Sie müssen außerdem sicherstellen, dass Amazon Redshift die IAM-Rolle während COPY- und UNLOAD-Operationen übernehmen kann.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Principal": {
        "Service": "redshift.amazonaws.com"
      },
      "Action": "sts:AssumeRole"
    }
  ]
}
```

Das folgende Beispiel verwendet die IAM-Authentifizierung zwischen Spark und Amazon Redshift:

```
from pyspark.sql import SQLContext
import boto3

sc = # existing SparkContext
sql_context = SQLContext(sc)

url = "jdbc:redshift:iam//redshift-host:redshift-port/db-name"
iam_role_arn = "arn:aws:iam::account-id:role/role-name"

# Read data from a table
df = sql_context.read \
    .format("io.github.spark_redshift_community.spark.redshift") \
    .option("url", url) \
    .option("aws_iam_role", iam_role_arn) \
    .option("dbtable", "my_table") \
    .option("tempdir", "s3a://path/for/temp/data") \
    .mode("error") \
    .load()
```



## Lesen und Schreiben von und zu Amazon Redshift

Die folgenden Codebeispiele werden verwendet PySpark , um Beispieldaten von und in eine Amazon Redshift Redshift-Datenbank mit Datenquellen-API und unter Verwendung von SparkSQL zu lesen und zu schreiben.

### Data source API

Wird verwendet PySpark , um Beispieldaten von und in eine Amazon Redshift Redshift-Datenbank mit Datenquellen-API zu lesen und zu schreiben.

```
import boto3
from pyspark.sql import SQLContext

sc = # existing SparkContext
sql_context = SQLContext(sc)

url = "jdbc:redshift:iam://redshifthost:5439/database"
aws_iam_role_arn = "arn:aws:iam::accountID:role/roleName"

df = sql_context.read \
    .format("io.github.spark_redshift_community.spark.redshift") \
    .option("url", url) \
    .option("dbtable", "tableName") \
    .option("tempdir", "s3://path/for/temp/data") \
    .option("aws_iam_role", "aws_iam_role_arn") \
    .load()

df.write \
    .format("io.github.spark_redshift_community.spark.redshift") \
    .option("url", url) \
    .option("dbtable", "tableName_copy") \
    .option("tempdir", "s3://path/for/temp/data") \
    .option("aws_iam_role", "aws_iam_role_arn") \
    .mode("error") \
    .save()
```

### SparkSQL

Wird verwendet PySpark , um Beispieldaten mit SparkSQL aus und in eine Amazon Redshift Redshift-Datenbank zu lesen und zu schreiben.

```
import boto3
import json
import sys
import os
from pyspark.sql import SparkSession

spark = SparkSession \
    .builder \
    .enableHiveSupport() \
    .getOrCreate()

url = "jdbc:redshift:iam://redshifthost:5439/database"
aws_iam_role_arn = "arn:aws:iam::accountID:role/roleName"

bucket = "s3://path/for/temp/data"
tableName = "tableName" # Redshift table name

s = f"""CREATE TABLE IF NOT EXISTS {tableName} (country string, data string)
    USING io.github.spark_redshift_community.spark.redshift
    OPTIONS (dbtable '{tableName}', tempdir '{bucket}', url '{url}', aws_iam_role
    '{aws_iam_role_arn'} '); """

spark.sql(s)

columns = ["country" ,"data"]
data = [("test-country", "test-data")]
df = spark.sparkContext.parallelize(data).toDF(columns)

# Insert data into table
df.write.insertInto(tableName, overwrite=False)
df = spark.sql(f"SELECT * FROM {tableName}")
df.show()
```

## Überlegungen und Einschränkungen bei der Verwendung des Spark-Connectors

- Wir empfehlen, SSL für die JDBC-Verbindung von Spark auf Amazon EMR zu Amazon Redshift zu aktivieren.

- Als bewährte Methode empfehlen wir Ihnen, die Anmeldeinformationen für den Amazon-Redshift-Cluster in AWS Secrets Manager zu verwalten. Ein Beispiel finden Sie [unter Verwenden AWS Secrets Manager zum Abrufen von Anmeldeinformationen für die Verbindung mit Amazon Redshift](#).
- Wir empfehlen, dass Sie eine IAM-Rolle mit dem Parameter `aws_iam_role` für den Amazon-Redshift-Authentifizierungsparameter zu übergeben.
- Derzeit wird das Parquet-Format vom Parameter `tempformat` nicht unterstützt.
- Die `tempdir`-URI verweist auf einen Amazon-S3-Speicherort. Dieses temporäre Verzeichnis wird nicht automatisch bereinigt und kann zusätzliche Kosten verursachen.
- Beachten Sie die folgenden Empfehlungen für Amazon Redshift:
  - Wir empfehlen, den öffentlichen Zugriff auf den Amazon-Redshift-Cluster zu blockieren.
  - Wir empfehlen, die [Amazon-Redshift-Auditprotokollierung](#) zu aktivieren.
  - Wir empfehlen Ihnen die [Amazon-Redshift-Verschlüsselung im Ruhezustand](#) zu aktivieren.
- Beachten Sie die folgenden Empfehlungen für Amazon S3:
  - Wir empfehlen Ihnen [den öffentlichen Zugriff auf Amazon-S3-Buckets zu blockieren](#).
  - Wir empfehlen die Verwendung der [serverseitigen Amazon-S3-Verschlüsselung](#), um die verwendeten Amazon-S3-Buckets zu verschlüsseln.
  - Wir empfehlen, die [Lebenszyklusrichtlinien für Amazon S3](#) zu verwenden, um die Aufbewahrungsregeln für den Amazon-S3-Bucket zu definieren.
  - Amazon EMR überprüft immer Code, der aus Open Source in das Image importiert wurde. Aus Sicherheitsgründen unterstützen wir die folgenden Authentifizierungsmethoden von Spark für Amazon S3 nicht:
    - AWS Zugriffstasten in der `hadoop-env` Konfigurationsklassifizierung einrichten
    - Kodierung der AWS Zugriffsschlüssel in der `tempdir` URI

Weitere Informationen zum Verwenden des Konnektors und seiner unterstützten Parameter finden Sie in den folgenden Ressourcen:

- [Amazon-Redshift-Integration für Apache Spark](#) im Amazon-Redshift-Verwaltungshandbuch
- Das [spark-redshift-Community-Repository](#) auf Github

## Spark-Versionsverlauf

In der folgenden Tabelle sind die Versionen von Spark aufgeführt, die in jeder Release-Version von Amazon EMR enthalten ist, zusammen mit den Komponenten, die mit der Anwendung installiert wurden. Informationen zu den Komponentenversionen in den einzelnen Versionen finden Sie im Abschnitt [Komponentenversion für Ihre Version in Amazon-EMR-7.x-Versionen](#), [Amazon-EMR-6.x-Versionen](#) oder [Amazon-EMR-5.x-Versionen](#).

### Important

Apache-Spark-Version 2.3.1, verfügbar ab Amazon-EMR-Version 5.16.0, adressiert [CVE-2018-8024](#) und [CVE-2018-1334](#). Wir empfehlen, dass Sie frühere Versionen von Spark zu Spark-Version 2.3.1 oder höher migrieren.

### Die Spark Versionsinformationen

| Amazon-EMR-Version<br>sbezeichnung | Spark-Version | Mit Spark installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                                         |
|------------------------------------|---------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.36.2                         | 2.4.8         | aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-7.1.0                          | 3.5.0         | delta, emrfs, emr-goodies, emr-ddb, emr-s3-select,                                                                                                                                                                                                                                                                                                                                            |

| Amazon-EMR-Version<br>sbezeichnung | Spark-Version | Mit Spark installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                               |
|------------------------------------|---------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                                    |               | hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave                                                    |
| emr-7.0.0                          | 3.5.0         | delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

| Amazon-EMR-Version<br>sbezeichnung | Spark-Version | Mit Spark installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                                                        |
|------------------------------------|---------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.15.0                         | 3.4.1         | aws-sagemaker-spark-sdk, delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-6.14.0                         | 3.4.1         | aws-sagemaker-spark-sdk, delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

| Amazon-EMR-Version<br>sbezeichnung | Spark-Version | Mit Spark installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                                                        |
|------------------------------------|---------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.13.0                         | 3.4.1         | aws-sagemaker-spark-sdk, delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-6.12.0                         | 3.4.0         | aws-sagemaker-spark-sdk, delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

| Amazon-EMR-Version<br>sbezeichnung | Spark-Version | Mit Spark installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                                                        |
|------------------------------------|---------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.11.1                         | 3.3.2         | aws-sagemaker-spark-sdk, delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-6.11.0                         | 3.3.2         | aws-sagemaker-spark-sdk, delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |



| Amazon-EMR-Version<br>sbezeichnung | Spark-Version | Mit Spark installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                                                                                                           |
|------------------------------------|---------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.10.1                         | 3.3.1         | aws-sagemaker-spark-sdk,<br>delta, emrfs, emr-goodies,<br>emr-ddb, emr-s3-select,<br>hadoop-client, hadoop-hdfs-<br>datanode, hadoop-hdfs-librar<br>y, hadoop-hdfs-namenode,<br>hadoop-httpfs-server, hadoop-<br>kms-server, hadoop-yarn-<br>nodemanager, hadoop-yarn-<br>resourcemanager, hadoop-ya<br>rn-timeline-server, hudi, hudi-<br>spark, iceberg, livy-server,<br>nginx, r, spark-client, spark-<br>history-server, spark-on-yarn,<br>spark-yarn-slave |
| emr-6.10.0                         | 3.3.1         | aws-sagemaker-spark-sdk,<br>delta, emrfs, emr-goodies,<br>emr-ddb, emr-s3-select,<br>hadoop-client, hadoop-hdfs-<br>datanode, hadoop-hdfs-librar<br>y, hadoop-hdfs-namenode,<br>hadoop-httpfs-server, hadoop-<br>kms-server, hadoop-yarn-<br>nodemanager, hadoop-yarn-<br>resourcemanager, hadoop-ya<br>rn-timeline-server, hudi, hudi-<br>spark, iceberg, livy-server,<br>nginx, r, spark-client, spark-<br>history-server, spark-on-yarn,<br>spark-yarn-slave |

| Amazon-EMR-Version<br>sbezeichnung | Spark-Version | Mit Spark installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                                                        |
|------------------------------------|---------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.9.1                          | 3.3.0         | aws-sagemaker-spark-sdk, delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-6.9.0                          | 3.3.0         | aws-sagemaker-spark-sdk, delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

| Amazon-EMR-Version<br>sbezeichnung | Spark-Version | Mit Spark installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                                                  |
|------------------------------------|---------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.8.1                          | 3.3.0         | aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-6.8.0                          | 3.3.0         | aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

| Amazon-EMR-Version<br>sbezeichnung | Spark-Version | Mit Spark installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                                                  |
|------------------------------------|---------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.7.0                          | 3.2.1         | aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-5.36.1                         | 2.4.8         | aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave          |

| Amazon-EMR-Version<br>sbezeichnung | Spark-Version | Mit Spark installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                                                  |
|------------------------------------|---------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.36.0                         | 2.4.8         | aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave          |
| emr-6.6.0                          | 3.2.0         | aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

| Amazon-EMR-Version<br>sbezeichnung | Spark-Version | Mit Spark installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                                                  |
|------------------------------------|---------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.35.0                         | 2.4.8         | aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave          |
| emr-6.5.0                          | 3.1.2         | aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

| Amazon-EMR-Version<br>sbezeichnung | Spark-Version | Mit Spark installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                                         |
|------------------------------------|---------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.4.0                          | 3.1.2         | aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-6.3.1                          | 3.1.1         | aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

| Amazon-EMR-Version<br>sbezeichnung | Spark-Version | Mit Spark installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                                         |
|------------------------------------|---------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.3.0                          | 3.1.1         | aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-6.2.1                          | 3.0.1         | aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |



| Amazon-EMR-Version<br>sbezeichnung | Spark-Version | Mit Spark installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                                         |
|------------------------------------|---------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.2.0                          | 3.0.1         | aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-6.1.1                          | 3.0.0         | aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

| Amazon-EMR-Version<br>sbezeichnung | Spark-Version | Mit Spark installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                                         |
|------------------------------------|---------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.1.0                          | 3.0.0         | aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-6.0.1                          | 2.4.4         | aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave             |

| Amazon-EMR-Version<br>sbezeichnung | Spark-Version | Mit Spark installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                                         |
|------------------------------------|---------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.0.0                          | 2.4.4         | aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave             |
| emr-5.34.0                         | 2.4.8         | aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

| Amazon-EMR-Version<br>sbezeichnung | Spark-Version | Mit Spark installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                                         |
|------------------------------------|---------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.33.1                         | 2.4.7         | aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-5.33.0                         | 2.4.7         | aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

| Amazon-EMR-Version<br>sbezeichnung | Spark-Version | Mit Spark installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                                         |
|------------------------------------|---------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.32.1                         | 2.4.7         | aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-5.32.0                         | 2.4.7         | aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

| Amazon-EMR-Version<br>sbezeichnung | Spark-Version | Mit Spark installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                                         |
|------------------------------------|---------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.31.1                         | 2.4.6         | aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-5.31.0                         | 2.4.6         | aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

| Amazon-EMR-Version<br>sbezeichnung | Spark-Version | Mit Spark installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                             |
|------------------------------------|---------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.30.2                         | 2.4.5         | aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-5.30.1                         | 2.4.5         | aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

| Amazon-EMR-Version<br>sbezeichnung | Spark-Version | Mit Spark installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                                              |
|------------------------------------|---------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.30.0                         | 2.4.5         | aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-notebook-env, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-5.29.0                         | 2.4.4         | aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave                   |



| Amazon-EMR-Version<br>sbezeichnung | Spark-Version | Mit Spark installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                             |
|------------------------------------|---------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.28.1                         | 2.4.4         | aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-5.28.0                         | 2.4.4         | aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

| Amazon-EMR-Version<br>sbezeichnung | Spark-Version | Mit Spark installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                       |
|------------------------------------|---------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.27.1                         | 2.4.4         | aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-5.27.0                         | 2.4.4         | aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

| Amazon-EMR-Version<br>sbezeichnung | Spark-Version | Mit Spark installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                                                                       |
|------------------------------------|---------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.26.0                         | 2.4.3         | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, emr-<br>ddb, emr-s3-select, hadoop-cl<br>ient, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, livy-server,<br>nginx, r, spark-client, spark-<br>history-server, spark-on-yarn,<br>spark-yarn-slave |
| emr-5.25.0                         | 2.4.3         | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, emr-<br>ddb, emr-s3-select, hadoop-cl<br>ient, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, livy-server,<br>nginx, r, spark-client, spark-<br>history-server, spark-on-yarn,<br>spark-yarn-slave |

| Amazon-EMR-Version<br>sbezeichnung | Spark-Version | Mit Spark installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                                                                       |
|------------------------------------|---------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.24.1                         | 2.4.2         | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, emr-<br>ddb, emr-s3-select, hadoop-cl<br>ient, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, livy-server,<br>nginx, r, spark-client, spark-<br>history-server, spark-on-yarn,<br>spark-yarn-slave |
| emr-5.24.0                         | 2.4.2         | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, emr-<br>ddb, emr-s3-select, hadoop-cl<br>ient, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, livy-server,<br>nginx, r, spark-client, spark-<br>history-server, spark-on-yarn,<br>spark-yarn-slave |

| Amazon-EMR-Version<br>sbezeichnung | Spark-Version | Mit Spark installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                       |
|------------------------------------|---------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.23.1                         | 2.4.0         | aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-5.23.0                         | 2.4.0         | aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

| Amazon-EMR-Version<br>sbezeichnung | Spark-Version | Mit Spark installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                       |
|------------------------------------|---------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.22.0                         | 2.4.0         | aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-5.21.2                         | 2.4.0         | aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

| Amazon-EMR-Version<br>sbezeichnung | Spark-Version | Mit Spark installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                       |
|------------------------------------|---------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.21.1                         | 2.4.0         | aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-5.21.0                         | 2.4.0         | aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

| Amazon-EMR-Version<br>sbezeichnung | Spark-Version | Mit Spark installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                       |
|------------------------------------|---------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.20.1                         | 2.4.0         | aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-5.20.0                         | 2.4.0         | aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |



| Amazon-EMR-Version<br>sbezeichnung | Spark-Version | Mit Spark installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                                                                       |
|------------------------------------|---------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.19.1                         | 2.3.2         | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, emr-<br>ddb, emr-s3-select, hadoop-cl<br>ient, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, livy-server,<br>nginx, r, spark-client, spark-<br>history-server, spark-on-yarn,<br>spark-yarn-slave |
| emr-5.19.0                         | 2.3.2         | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, emr-<br>ddb, emr-s3-select, hadoop-cl<br>ient, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, livy-server,<br>nginx, r, spark-client, spark-<br>history-server, spark-on-yarn,<br>spark-yarn-slave |

| Amazon-EMR-Version<br>sbezeichnung | Spark-Version | Mit Spark installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                                                                       |
|------------------------------------|---------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.18.1                         | 2.3.2         | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, emr-<br>ddb, emr-s3-select, hadoop-cl<br>ient, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, livy-server,<br>nginx, r, spark-client, spark-<br>history-server, spark-on-yarn,<br>spark-yarn-slave |
| emr-5.18.0                         | 2.3.2         | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, emr-<br>ddb, emr-s3-select, hadoop-cl<br>ient, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, livy-server,<br>nginx, r, spark-client, spark-<br>history-server, spark-on-yarn,<br>spark-yarn-slave |

| Amazon-EMR-Version<br>sbezeichnung | Spark-Version | Mit Spark installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                |
|------------------------------------|---------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.17.2                         | 2.3.1         | aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-5.17.1                         | 2.3.1         | aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

| Amazon-EMR-Version<br>sbezeichnung | Spark-Version | Mit Spark installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                |
|------------------------------------|---------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.17.0                         | 2.3.1         | aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-5.16.1                         | 2.3.1         | aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave                |

| Amazon-EMR-Version<br>sbezeichnung | Spark-Version | Mit Spark installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                                            |
|------------------------------------|---------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.16.0                         | 2.3.1         | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, emr-ddb,<br>hadoop-client, hadoop-hdfs-<br>datanode, hadoop-hdfs-librar<br>y, hadoop-hdfs-namenode,<br>hadoop-httfs-server, hadoop-<br>kms-server, hadoop-yarn-<br>nodemanager, hadoop-yarn-<br>resourcemanager, hadoop-ya<br>rn-timeline-server, livy-server,<br>r, spark-client, spark-history-<br>server, spark-on-yarn, spark-<br>yarn-slave |
| emr-5.15.1                         | 2.3.0         | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, emr-ddb,<br>hadoop-client, hadoop-hdfs-<br>datanode, hadoop-hdfs-librar<br>y, hadoop-hdfs-namenode,<br>hadoop-httfs-server, hadoop-<br>kms-server, hadoop-yarn-<br>nodemanager, hadoop-yarn-<br>resourcemanager, hadoop-ya<br>rn-timeline-server, livy-server,<br>r, spark-client, spark-history-<br>server, spark-on-yarn, spark-<br>yarn-slave |

| Amazon-EMR-Version<br>sbezeichnung | Spark-Version | Mit Spark installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                                            |
|------------------------------------|---------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.15.0                         | 2.3.0         | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, emr-ddb,<br>hadoop-client, hadoop-hdfs-<br>datanode, hadoop-hdfs-librar<br>y, hadoop-hdfs-namenode,<br>hadoop-httfs-server, hadoop-<br>kms-server, hadoop-yarn-<br>nodemanager, hadoop-yarn-<br>resourcemanager, hadoop-ya<br>rn-timeline-server, livy-server,<br>r, spark-client, spark-history-<br>server, spark-on-yarn, spark-<br>yarn-slave |
| emr-5.14.2                         | 2.3.0         | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, emr-ddb,<br>hadoop-client, hadoop-hdfs-<br>datanode, hadoop-hdfs-librar<br>y, hadoop-hdfs-namenode,<br>hadoop-httfs-server, hadoop-<br>kms-server, hadoop-yarn-<br>nodemanager, hadoop-yarn-<br>resourcemanager, hadoop-ya<br>rn-timeline-server, livy-server,<br>r, spark-client, spark-history-<br>server, spark-on-yarn, spark-<br>yarn-slave |

| Amazon-EMR-Version<br>sbezeichnung | Spark-Version | Mit Spark installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                                            |
|------------------------------------|---------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.14.1                         | 2.3.0         | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, emr-ddb,<br>hadoop-client, hadoop-hdfs-<br>datanode, hadoop-hdfs-librar<br>y, hadoop-hdfs-namenode,<br>hadoop-httfs-server, hadoop-<br>kms-server, hadoop-yarn-<br>nodemanager, hadoop-yarn-<br>resourcemanager, hadoop-ya<br>rn-timeline-server, livy-server,<br>r, spark-client, spark-history-<br>server, spark-on-yarn, spark-<br>yarn-slave |
| emr-5.14.0                         | 2.3.0         | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, emr-ddb,<br>hadoop-client, hadoop-hdfs-<br>datanode, hadoop-hdfs-librar<br>y, hadoop-hdfs-namenode,<br>hadoop-httfs-server, hadoop-<br>kms-server, hadoop-yarn-<br>nodemanager, hadoop-yarn-<br>resourcemanager, hadoop-ya<br>rn-timeline-server, livy-server,<br>r, spark-client, spark-history-<br>server, spark-on-yarn, spark-<br>yarn-slave |

| Amazon-EMR-Version<br>sbezeichnung | Spark-Version | Mit Spark installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                                            |
|------------------------------------|---------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.13.1                         | 2.3.0         | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, emr-ddb,<br>hadoop-client, hadoop-hdfs-<br>datanode, hadoop-hdfs-librar<br>y, hadoop-hdfs-namenode,<br>hadoop-httfs-server, hadoop-<br>kms-server, hadoop-yarn-<br>nodemanager, hadoop-yarn-<br>resourcemanager, hadoop-ya<br>rn-timeline-server, livy-server,<br>r, spark-client, spark-history-<br>server, spark-on-yarn, spark-<br>yarn-slave |
| emr-5.13.0                         | 2.3.0         | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, emr-ddb,<br>hadoop-client, hadoop-hdfs-<br>datanode, hadoop-hdfs-librar<br>y, hadoop-hdfs-namenode,<br>hadoop-httfs-server, hadoop-<br>kms-server, hadoop-yarn-<br>nodemanager, hadoop-yarn-<br>resourcemanager, hadoop-ya<br>rn-timeline-server, livy-server,<br>r, spark-client, spark-history-<br>server, spark-on-yarn, spark-<br>yarn-slave |



| Amazon-EMR-Version<br>sbezeichnung | Spark-Version | Mit Spark installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                                         |
|------------------------------------|---------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.12.3                         | 2.2.1         | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, emr-ddb,<br>hadoop-client, hadoop-hdfs-<br>datanode, hadoop-hdfs-librar<br>y, hadoop-hdfs-namenode,<br>hadoop-httfs-server, hadoop-<br>kms-server, hadoop-yarn-<br>nodemanager, hadoop-yarn-<br>resourcemanager, hadoop-ya<br>rn-timeline-server, livy-server,<br>spark-client, spark-history-serv<br>er, spark-on-yarn, spark-yarn-<br>slave |
| emr-5.12.2                         | 2.2.1         | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, emr-ddb,<br>hadoop-client, hadoop-hdfs-<br>datanode, hadoop-hdfs-librar<br>y, hadoop-hdfs-namenode,<br>hadoop-httfs-server, hadoop-<br>kms-server, hadoop-yarn-<br>nodemanager, hadoop-yarn-<br>resourcemanager, hadoop-ya<br>rn-timeline-server, livy-server,<br>spark-client, spark-history-serv<br>er, spark-on-yarn, spark-yarn-<br>slave |

| Amazon-EMR-Version<br>sbezeichnung | Spark-Version | Mit Spark installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                                         |
|------------------------------------|---------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.12.1                         | 2.2.1         | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, emr-ddb,<br>hadoop-client, hadoop-hdfs-<br>datanode, hadoop-hdfs-librar<br>y, hadoop-hdfs-namenode,<br>hadoop-httfs-server, hadoop-<br>kms-server, hadoop-yarn-<br>nodemanager, hadoop-yarn-<br>resourcemanager, hadoop-ya<br>rn-timeline-server, livy-server,<br>spark-client, spark-history-serv<br>er, spark-on-yarn, spark-yarn-<br>slave |
| emr-5.12.0                         | 2.2.1         | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, emr-ddb,<br>hadoop-client, hadoop-hdfs-<br>datanode, hadoop-hdfs-librar<br>y, hadoop-hdfs-namenode,<br>hadoop-httfs-server, hadoop-<br>kms-server, hadoop-yarn-<br>nodemanager, hadoop-yarn-<br>resourcemanager, hadoop-ya<br>rn-timeline-server, livy-server,<br>spark-client, spark-history-serv<br>er, spark-on-yarn, spark-yarn-<br>slave |

| Amazon-EMR-Version<br>sbezeichnung | Spark-Version | Mit Spark installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                                         |
|------------------------------------|---------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.11.4                         | 2.2.1         | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, emr-ddb,<br>hadoop-client, hadoop-hdfs-<br>datanode, hadoop-hdfs-librar<br>y, hadoop-hdfs-namenode,<br>hadoop-httfs-server, hadoop-<br>kms-server, hadoop-yarn-<br>nodemanager, hadoop-yarn-<br>resourcemanager, hadoop-ya<br>rn-timeline-server, livy-server,<br>spark-client, spark-history-serv<br>er, spark-on-yarn, spark-yarn-<br>slave |
| emr-5.11.3                         | 2.2.1         | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, emr-ddb,<br>hadoop-client, hadoop-hdfs-<br>datanode, hadoop-hdfs-librar<br>y, hadoop-hdfs-namenode,<br>hadoop-httfs-server, hadoop-<br>kms-server, hadoop-yarn-<br>nodemanager, hadoop-yarn-<br>resourcemanager, hadoop-ya<br>rn-timeline-server, livy-server,<br>spark-client, spark-history-serv<br>er, spark-on-yarn, spark-yarn-<br>slave |

| Amazon-EMR-Version<br>sbezeichnung | Spark-Version | Mit Spark installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                                         |
|------------------------------------|---------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.11.2                         | 2.2.1         | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, emr-ddb,<br>hadoop-client, hadoop-hdfs-<br>datanode, hadoop-hdfs-librar<br>y, hadoop-hdfs-namenode,<br>hadoop-httfs-server, hadoop-<br>kms-server, hadoop-yarn-<br>nodemanager, hadoop-yarn-<br>resourcemanager, hadoop-ya<br>rn-timeline-server, livy-server,<br>spark-client, spark-history-serv<br>er, spark-on-yarn, spark-yarn-<br>slave |
| emr-5.11.1                         | 2.2.1         | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, emr-ddb,<br>hadoop-client, hadoop-hdfs-<br>datanode, hadoop-hdfs-librar<br>y, hadoop-hdfs-namenode,<br>hadoop-httfs-server, hadoop-<br>kms-server, hadoop-yarn-<br>nodemanager, hadoop-yarn-<br>resourcemanager, hadoop-ya<br>rn-timeline-server, livy-server,<br>spark-client, spark-history-serv<br>er, spark-on-yarn, spark-yarn-<br>slave |

| Amazon-EMR-Version<br>sbezeichnung | Spark-Version | Mit Spark installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                                          |
|------------------------------------|---------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.11.0                         | 2.2.1         | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, emr-ddb,<br>hadoop-client, hadoop-hdfs-<br>datanode, hadoop-hdfs-librar<br>y, hadoop-hdfs-namenode,<br>hadoop-httpfs-server, hadoop-<br>kms-server, hadoop-yarn-<br>nodemanager, hadoop-yarn-<br>resourcemanager, hadoop-ya<br>rn-timeline-server, livy-server,<br>spark-client, spark-history-serv<br>er, spark-on-yarn, spark-yarn-<br>slave |
| emr-5.10.1                         | 2.2.0         | emrfs, emr-goodies, emr-ddb,<br>hadoop-client, hadoop-hdfs-<br>datanode, hadoop-hdfs-librar<br>y, hadoop-hdfs-namenode,<br>hadoop-httpfs-server, hadoop-<br>kms-server, hadoop-yarn-<br>nodemanager, hadoop-yarn-<br>resourcemanager, hadoop-ya<br>rn-timeline-server, livy-server,<br>spark-client, spark-history-serv<br>er, spark-on-yarn, spark-yarn-<br>slave                             |

| Amazon-EMR-Version<br>sbezeichnung | Spark-Version | Mit Spark installierte<br>Komponenten                                                                                                                                                                                                                                                                                    |
|------------------------------------|---------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.10.0                         | 2.2.0         | emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-5.9.1                          | 2.2.0         | emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave              |

| Amazon-EMR-Version<br>sbezeichnung | Spark-Version | Mit Spark installierte<br>Komponenten                                                                                                                                                                                                                                                                       |
|------------------------------------|---------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.9.0                          | 2.2.0         | emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-5.8.3                          | 2.2.0         | emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

| Amazon-EMR-Version<br>sbezeichnung | Spark-Version | Mit Spark installierte<br>Komponenten                                                                                                                                                                                                                                                                       |
|------------------------------------|---------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.8.2                          | 2.2.0         | emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-5.8.1                          | 2.2.0         | emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |



| Amazon-EMR-Version<br>sbezeichnung | Spark-Version | Mit Spark installierte<br>Komponenten                                                                                                                                                                                                                                                                       |
|------------------------------------|---------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.8.0                          | 2.2.0         | emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-5.7.1                          | 2.1.1         | emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

| Amazon-EMR-Version<br>sbezeichnung | Spark-Version | Mit Spark installierte<br>Komponenten                                                                                                                                                                                                                                                                       |
|------------------------------------|---------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.7.0                          | 2.1.1         | emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-5.6.1                          | 2.1.1         | emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

| Amazon-EMR-Version<br>sbezeichnung | Spark-Version | Mit Spark installierte<br>Komponenten                                                                                                                                                                                                                                                                       |
|------------------------------------|---------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.6.0                          | 2.1.1         | emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-5.5.4                          | 2.1.0         | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave                                       |
| emr-5.5.3                          | 2.1.0         | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave                                       |

| Amazon-EMR-Version<br>sbezeichnung | Spark-Version | Mit Spark installierte<br>Komponenten                                                                                                                                                                                                                                  |
|------------------------------------|---------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.5.2                          | 2.1.0         | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-5.5.1                          | 2.1.0         | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-5.5.0                          | 2.1.0         | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

| Amazon-EMR-Version<br>sbezeichnung | Spark-Version | Mit Spark installierte<br>Komponenten                                                                                                                                                                                                                                 |
|------------------------------------|---------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.4.1                          | 2.1.0         | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-5.4.0                          | 2.1.0         | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-5.3.2                          | 2.1.0         | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

| Amazon-EMR-Version<br>sbezeichnung | Spark-Version | Mit Spark installierte<br>Komponenten                                                                                                                                                                                                                                  |
|------------------------------------|---------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.3.1                          | 2.1.0         | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-5.3.0                          | 2.1.0         | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-5.2.3                          | 2.0.2         | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

| Amazon-EMR-Version<br>sbezeichnung | Spark-Version | Mit Spark installierte<br>Komponenten                                                                                                                                                                                                                                  |
|------------------------------------|---------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.2.2                          | 2.0.2         | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-5.2.1                          | 2.0.2         | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-5.2.0                          | 2.0.2         | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

| Amazon-EMR-Version<br>sbezeichnung | Spark-Version | Mit Spark installierte<br>Komponenten                                                                                                                                                                                                                                  |
|------------------------------------|---------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.1.1                          | 2.0.1         | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-5.1.0                          | 2.0.1         | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-5.0.3                          | 2.0.1         | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |



| Amazon-EMR-Version<br>sbezeichnung | Spark-Version | Mit Spark installierte<br>Komponenten                                                                                                                                                                                                                                  |
|------------------------------------|---------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.0.2                          | 2.0.0         | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-5.0.1                          | 2.0.0         | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-5.0.0                          | 2.0.0         | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

| Amazon-EMR-Version<br>sbezeichnung | Spark-Version | Mit Spark installierte<br>Komponenten                                                                                                                                                                                                                                  |
|------------------------------------|---------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-4.9.6                          | 1.6.3         | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-4.9.5                          | 1.6.3         | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-4.9.4                          | 1.6.3         | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

| Amazon-EMR-Version<br>sbezeichnung | Spark-Version | Mit Spark installierte<br>Komponenten                                                                                                                                                                                                                                  |
|------------------------------------|---------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-4.9.3                          | 1.6.3         | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-4.9.2                          | 1.6.3         | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-4.9.1                          | 1.6.3         | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

| Amazon-EMR-Version<br>sbezeichnung | Spark-Version | Mit Spark installierte<br>Komponenten                                                                                                                                                                                                                                  |
|------------------------------------|---------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-4.8.5                          | 1.6.3         | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-4.8.4                          | 1.6.3         | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-4.8.3                          | 1.6.3         | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

| Amazon-EMR-Version<br>sbezeichnung | Spark-Version | Mit Spark installierte<br>Komponenten                                                                                                                                                                                                                                 |
|------------------------------------|---------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-4.8.2                          | 1.6.2         | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-4.8.1                          | 1.6.2         | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-4.8.0                          | 1.6.2         | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

| Amazon-EMR-Version<br>sbezeichnung | Spark-Version | Mit Spark installierte<br>Komponenten                                                                                                                                                                                                                                  |
|------------------------------------|---------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-4.7.4                          | 1.6.2         | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-4.7.3                          | 1.6.2         | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-4.7.2                          | 1.6.2         | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

| Amazon-EMR-Version<br>sbezeichnung | Spark-Version | Mit Spark installierte<br>Komponenten                                                                                                                                                                                                                                 |
|------------------------------------|---------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-4.7.1                          | 1.6.1         | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-4.7.0                          | 1.6.1         | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-4.6.1                          | 1.6.1         | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave                    |

| Amazon-EMR-Version<br>sbezeichnung | Spark-Version | Mit Spark installierte<br>Komponenten                                                                                                                                                                                                                       |
|------------------------------------|---------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-4.6.0                          | 1.6.1         | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ht<br>tpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resource<br>manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-4.5.0                          | 1.6.1         | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ht<br>tpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resource<br>manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-4.4.0                          | 1.6.0         | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ht<br>tpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resource<br>manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |



| Amazon-EMR-Version<br>sbezeichnung | Spark-Version | Mit Spark installierte<br>Komponenten                                                                                                                                                                                                                       |
|------------------------------------|---------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-4.3.0                          | 1.6.0         | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ht<br>tpfs-server, hadoop-yarn-nodemanager, hadoop-ya<br>rn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-4.2.0                          | 1.5.2         | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ht<br>tpfs-server, hadoop-yarn-nodemanager, hadoop-ya<br>rn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-4.1.0                          | 1.5.0         | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ht<br>tpfs-server, hadoop-yarn-nodemanager, hadoop-ya<br>rn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

| Amazon-EMR-Version<br>sbezeichnung | Spark-Version | Mit Spark installierte<br>Komponenten                                                                                                                                                                                                                   |
|------------------------------------|---------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-4.0.0                          | 1.4.1         | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datano<br>de, hadoop-hdfs-namenode,<br>hadoop-ftpfs-server, hadoop-<br>yarn-nodemanager, hadoop-<br>yarn-resourcemanager, spark-<br>client, spark-history-server,<br>spark-on-yarn, spark-yarn-<br>slave |

# Apache Sqoop

Apache Sqoop ist ein Tool zur Übertragung von Daten zwischen Amazon-S3-, Hadoop-, HDFS- und RDBMS-Datenbanken. Weitere Informationen finden Sie auf der [Apache Sqoop-Website](#). Sqoop ist in Amazon-EMR-Versionen 5.0.0 und höher enthalten. Frühere Versionen enthalten Sqoop als Sandbox-Anwendung. Weitere Informationen finden Sie unter [Amazon-EMR-4.x-Versionen](#).

## Themen

- [Die Sqoop-Versionsinformationen](#)
- [Überlegungen zu Sqoop auf Amazon EMR](#)
- [Versionsverlauf von Sqoop](#)

## Die Sqoop-Versionsinformationen

### Sqoop Version für 7.1.0

Die folgende Tabelle listet die Version von Sqoop auf, die in der neuesten Version der Amazon-EMR-7.x-Serie enthalten ist, zusammen mit den Komponenten, die Amazon EMR mit Sqoop installiert.

Die Version der Komponenten, die in dieser Version mit Sqoop installiert wurden, finden Sie unter [Komponentenversionen von Version 7.1.0](#).

### Sqoop Versionsinformationen für emr-7.1.0

| Amazon-EMR-Version<br>sbezeichnung | Sqoop-Version | Mit Sqoop installierte<br>Komponenten                                                                                                                                                                                                   |
|------------------------------------|---------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-7.1.0                          | Sqoop 1.4.7   | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn- |

| Amazon-EMR-Version<br>sbezeichnung | Sqoop-Version | Mit Sqoop installierte<br>Komponenten         |
|------------------------------------|---------------|-----------------------------------------------|
|                                    |               | timeline-server, mariadb-server, sqoop-client |

### Sqoop-Version für 6.15.0

In der folgenden Tabelle wird die Version von Sqoop, die in der neuesten Version von Amazon-EMR-6.x-Serie enthalten ist, zusammen mit den Komponenten aufgeführt, die Amazon EMR mit Sqoop installiert.

Die Version der Komponenten, die mit Sqoop in dieser Version installiert wurden, finden Sie unter [Komponentenversionen der Version 6.15.0](#).

### Sqoop-Versionsinformationen für emr-6.15.0

| Amazon-EMR-Version<br>sbezeichnung | Sqoop-Version | Mit Sqoop installierte<br>Komponenten                                                                                                                                                                                                                                               |
|------------------------------------|---------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.15.0                         | Sqoop 1.4.7   | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client |

### Sqoop Version für 5.36.2

Die folgende Tabelle listet die Version von Sqoop auf, die in der neuesten Version der Amazon-EMR-5.x-Serie enthalten ist, zusammen mit den Komponenten, die Amazon EMR mit Sqoop installiert.

[Die Version der Komponenten, die in dieser Version mit Sqoop installiert wurden, finden Sie unter Komponentenversionen von Version 5.36.2.](#)

Sqoop Versionsinformationen für emr-5.36.2

| Amazon-EMR-Version<br>sbezeichnung | Sqoop-Version | Mit Sqoop installierte<br>Komponenten                                                                                                                                                                                                                                               |
|------------------------------------|---------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.36.2                         | Sqoop 1.4.7   | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client |

## Überlegungen zu Sqoop auf Amazon EMR

Beachten Sie folgende Punkte, wenn Sie Sqoop auf Amazon EMR ausführen.

### Sqoop mit der HCatalog-Integration verwenden

Sqoop auf Amazon EMR unterstützt die [Sqoop-HCatalog-Integration](#). Wenn Sie Sqoop verwenden, um Ausgaben in eine HCatalog-Tabelle in Amazon S3 zu schreiben, deaktivieren Sie Amazon-EMR-Direct-Write, indem Sie die Eigenschaften `mapred.output.direct.NativeS3FileSystem` und `mapred.output.direct.EmrFileSystem` auf `false` setzen. Weitere Informationen finden Sie unter [Verwenden von HCatalog](#). Sie können die Hadoop-Befehle `-D mapred.output.direct.NativeS3FileSystem=false` und `-D mapred.output.direct.EmrFileSystem=false` verwenden. Wenn Sie die Direct-Write-Funktion nicht deaktivieren, tritt kein Fehler auf, aber die Tabelle wird in Amazon S3 erstellt und es werden keine Daten geschrieben.

## Sqoop JDBC- und Datenbank-Support

Standardmäßig ist für Sqoop eine MariaDB und ein PostgreSQL-Treiber installiert. Der für Sqoop installierte PostgreSQL-Treiber funktioniert nur für PostgreSQL 8.4. Wenn Sie andere JDBC-Konnektoren für Sqoop installieren möchten, stellen Sie eine Verbindung mit dem Master-Knoten des Clusters her, und installieren Sie die Konnektoren in `/usr/lib/sqoop/lib`. Die folgenden Links führen zu verschiedenen JDBC-Konnektoren:

- MariaDB: [Informationen zu MariaDB Connector/J](#).
- PostgreSQL: [PostgreSQL-JDBC-Treiber](#).
- SQLServer: [Laden Sie den Microsoft-JDBC-Treiber für SQL Server herunter](#).
- MySQL: [Laden Sie Connector/J herunter](#).
- Oracle: [Rufen Sie die Oracle JDBC-Treiber und UCP aus dem Maven-Repository von Oracle ab](#)

*Die unterstützten Datenbanken für Sqoop sind unter der folgenden URL aufgeführt: [http://sqoop.apache.org/docs/ version/SqoopUserGuide.html#\\_supported\\_databases](http://sqoop.apache.org/docs/version/SqoopUserGuide.html#_supported_databases), wobei Version die Version von Sqoop ist, die Sie verwenden, zum Beispiel 1.4.6. Wenn die JDBC-Verbindungszeichenfolge nicht in der Liste aufgeführt ist, müssen Sie einen Treiber angeben.*

Sie können beispielsweise Daten in eine Amazon Redshift Datenbanktabelle mit dem folgenden Befehl (für JDBC-4.1) exportieren:

```
sqoop export --connect jdbc:redshift://$MYREDSHIFTHOST:5439/mydb --table mysqoopexport
--export-dir s3://mybucket/myinputfiles/ --driver com.amazon.redshift.jdbc41.Driver --
username master --password Mymasterpass1
```

Sie können sowohl die MariaDB- als auch die MySQL-Verbindungszeichenfolgen verwenden. Wenn Sie jedoch die MariaDB-Verbindungszeichenfolge angeben, müssen Sie den Treiber festlegen:

```
sqoop export --connect jdbc:mariadb://$HOSTNAME:3306/mydb --table mysqoopexport
--export-dir s3://mybucket/myinputfiles/ --driver org.mariadb.jdbc.Driver --
username master --password Mymasterpass1
```

Wenn Sie die Secure Socket Layer-Verschlüsselung für den Zugriff auf Ihre Datenbank verwenden, müssen Sie eine JDBC-URI wie im folgenden Sqoop-Exportbeispiel nutzen:

```
sqoop export --connect jdbc:mariadb://$HOSTNAME:3306/mydb?  
verifyServerCertificate=false&useSSL=true&requireSSL=true --table mysqoopexport  
--export-dir s3://mybucket/myinputfiles/ --driver org.mariadb.jdbc.Driver --  
username master --password Mymasterpass1
```

Weitere Informationen über die SSL-Verschlüsselung in RDS finden Sie unter [Verwenden von SSL zum Verschlüsseln einer Verbindung zu einer DB-Instance](#) im Amazon-RDS-Benutzerhandbuch.

Weitere Informationen finden Sie in der [Apache Sqoop](#)-Dokumentation.

## Schützen Ihres Passworts

Es gibt mehrere Methoden, wie Sie Ihr Passwort sicher weitergeben können:

### Java KeyStore

Die bevorzugte Methode verschlüsselt das Passwort mit Java KeyStore (JKS), sodass das Passwort nicht in einem lesbaren Format gespeichert werden muss.

1. Erstellen Sie einen Passwort-Alias. Geben Sie bei entsprechender Aufforderung das Passwort ein, mit dem Sie auf die Datenbank zugreifen.

```
hadoop credential create mydb.password.alias -provider jceks://hdfs/user/root/  
mysql.password.jceks
```

2. Verwenden Sie den Passwort-Alias, um den Sqoop-Auftrag zu starten:

```
sqoop export -Dhadoop.security.credential.provider.path=jceks://hdfs/user/  
root/mysql.password.jceks --connect jdbc:mariadb://$HOSTNAME:3306/mydb  
--table mysqoopexport --export-dir s3://mybucket/myinputfiles/ --driver  
org.mariadb.jdbc.Driver --username master --password-alias mydb.password.alias
```

### --password-file

Sie können wie im folgenden Beispiel dargestellt den `--password-file`-Befehl verwenden, um das Passwort durch eine Datei zu übergeben:

1. Erstellen Sie eine neue Datei, die das Passwort enthält:

```
echo -n 'Mymasterpass1' > /home/hadoop/mysql-pass.password
```

## 2. Verwenden Sie die Datei, um den Sqoop-Auftrag zu starten:

```
sqoop export --connect jdbc:mariadb://$HOSTNAME:3306/mydb --table mysqoopexport
--export-dir s3://mybucket/myinputfiles/ --driver org.mariadb.jdbc.Driver --
username master --password-file /home/hadoop/mysql-pass.password
```

-P

Sie können wie im folgenden Beispiel dargestellt den -P-Befehl verwenden, um das Passwort durch eine Aufforderung zu übergeben:

```
sqoop export --connect jdbc:mariadb://$HOSTNAME:3306/mydb --table mysqoopexport --
export-dir s3://mybucket/myinputfiles/ --driver org.mariadb.jdbc.Driver --username
master -P
```

## Versionsverlauf von Sqoop

In der folgenden Tabelle sind die Version von Sqoop aufgeführt, die in jeder Release-Version von Amazon EMR enthalten ist, zusammen mit den Komponenten, die mit der Anwendung installiert wurden. Informationen zu den Komponentenversionen in den einzelnen Versionen finden Sie im Abschnitt [Komponentenversion für Ihre Version in Amazon-EMR-7.x-Versionen](#), [Amazon-EMR-6.x-Versionen](#) oder [Amazon-EMR-5.x-Versionen](#).

Die Sqoop-Versionsinformationen

| Amazon-EMR-Version<br>sbezeichnung | Sqoop-Version | Mit Sqoop installierte<br>Komponenten                                                                                                                                                                                                   |
|------------------------------------|---------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.36.2                         | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn- |



| Amazon-EMR-Version<br>sbezeichnung | Sqoop-Version | Mit Sqoop installierte<br>Komponenten                                                                                                                                                                                                                                               |
|------------------------------------|---------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                                    |               | timeline-server, mariadb-server, sqoop-client                                                                                                                                                                                                                                       |
| emr-7.1.0                          | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client |
| emr-7.0.0                          | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client |

| Amazon-EMR-Version<br>sbezeichnung | Sqoop-Version | Mit Sqoop installierte<br>Komponenten                                                                                                                                                                                                                                               |
|------------------------------------|---------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.15.0                         | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client |
| emr-6.14.0                         | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client |

| Amazon-EMR-Version<br>sbezeichnung | Sqoop-Version | Mit Sqoop installierte<br>Komponenten                                                                                                                                                                                                                                               |
|------------------------------------|---------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.13.0                         | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client |
| emr-6.12.0                         | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client |

| Amazon-EMR-Version<br>sbezeichnung | Sqoop-Version | Mit Sqoop installierte<br>Komponenten                                                                                                                                                                                                                                               |
|------------------------------------|---------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.11.1                         | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client |
| emr-6.11.0                         | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client |

| Amazon-EMR-Version<br>sbezeichnung | Sqoop-Version | Mit Sqoop installierte<br>Komponenten                                                                                                                                                                                                                                                |
|------------------------------------|---------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.10.1                         | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mahout, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client |
| emr-6.10.0                         | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mahout, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client |

| Amazon-EMR-Version<br>sbezeichnung | Sqoop-Version | Mit Sqoop installierte<br>Komponenten                                                                                                                                                                                                                                               |
|------------------------------------|---------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.9.1                          | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client |
| emr-6.9.0                          | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client |

| Amazon-EMR-Version<br>sbezeichnung | Sqoop-Version | Mit Sqoop installierte<br>Komponenten                                                                                                                                                                                                                                               |
|------------------------------------|---------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.8.1                          | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client |
| emr-6.8.0                          | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client |

| Amazon-EMR-Version<br>sbezeichnung | Sqoop-Version | Mit Sqoop installierte<br>Komponenten                                                                                                                                                                                                                                               |
|------------------------------------|---------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.7.0                          | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client |
| emr-5.36.1                         | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client |



| Amazon-EMR-Version<br>sbezeichnung | Sqoop-Version | Mit Sqoop installierte<br>Komponenten                                                                                                                                                                                                                                               |
|------------------------------------|---------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.36.0                         | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client |
| emr-6.6.0                          | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client |

| Amazon-EMR-Version<br>sbezeichnung | Sqoop-Version | Mit Sqoop installierte<br>Komponenten                                                                                                                                                                                                                                               |
|------------------------------------|---------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.35.0                         | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client |
| emr-6.5.0                          | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client |

| Amazon-EMR-Version<br>sbezeichnung | Sqoop-Version | Mit Sqoop installierte<br>Komponenten                                                                                                                                                                                                                                               |
|------------------------------------|---------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.4.0                          | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client |
| emr-6.3.1                          | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client |

| Amazon-EMR-Version<br>sbezeichnung | Sqoop-Version | Mit Sqoop installierte<br>Komponenten                                                                                                                                                                                                                                               |
|------------------------------------|---------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.3.0                          | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client |
| emr-6.2.1                          | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client |

| Amazon-EMR-Version<br>sbezeichnung | Sqoop-Version | Mit Sqoop installierte<br>Komponenten                                                                                                                                                                                                                                               |
|------------------------------------|---------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.2.0                          | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client |
| emr-6.1.1                          | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client |

| Amazon-EMR-Version<br>sbezeichnung | Sqoop-Version | Mit Sqoop installierte<br>Komponenten                                                                                                                                                                                                                                               |
|------------------------------------|---------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.1.0                          | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client |
| emr-5.34.0                         | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client |

| Amazon-EMR-Version<br>sbezeichnung | Sqoop-Version | Mit Sqoop installierte<br>Komponenten                                                                                                                                                                                                                                               |
|------------------------------------|---------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.33.1                         | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client |
| emr-5.33.0                         | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client |

| Amazon-EMR-Version<br>sbezeichnung | Sqoop-Version | Mit Sqoop installierte<br>Komponenten                                                                                                                                                                                                                                               |
|------------------------------------|---------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.32.1                         | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client |
| emr-5.32.0                         | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client |



| Amazon-EMR-Version<br>sbezeichnung | Sqoop-Version | Mit Sqoop installierte<br>Komponenten                                                                                                                                                                                                                                               |
|------------------------------------|---------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.31.1                         | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client |
| emr-5.31.0                         | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client |

| Amazon-EMR-Version<br>sbezeichnung | Sqoop-Version | Mit Sqoop installierte<br>Komponenten                                                                                                                                                                                                                                                |
|------------------------------------|---------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.30.2                         | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mahout, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client |
| emr-5.30.1                         | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mahout, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client |

| Amazon-EMR-Version<br>sbezeichnung | Sqoop-Version | Mit Sqoop installierte<br>Komponenten                                                                                                                                                                                                                                                |
|------------------------------------|---------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.30.0                         | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client |
| emr-5.29.0                         | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client   |

| Amazon-EMR-Version<br>sbezeichnung | Sqoop-Version | Mit Sqoop installierte<br>Komponenten                                                                                                                                                                                                                                             |
|------------------------------------|---------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.28.1                         | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client |
| emr-5.28.0                         | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client |

| Amazon-EMR-Version<br>sbezeichnung | Sqoop-Version | Mit Sqoop installierte<br>Komponenten                                                                                                                                                                                                                                             |
|------------------------------------|---------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.27.1                         | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client |
| emr-5.27.0                         | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client |

| Amazon-EMR-Version<br>sbezeichnung | Sqoop-Version | Mit Sqoop installierte<br>Komponenten                                                                                                                                                                                                                                             |
|------------------------------------|---------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.26.0                         | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client |
| emr-5.25.0                         | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client |

| Amazon-EMR-Version<br>sbezeichnung | Sqoop-Version | Mit Sqoop installierte<br>Komponenten                                                                                                                                                                                                                                             |
|------------------------------------|---------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.24.1                         | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client |
| emr-5.24.0                         | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client |

| Amazon-EMR-Version<br>sbezeichnung | Sqoop-Version | Mit Sqoop installierte<br>Komponenten                                                                                                                                                                                                                                             |
|------------------------------------|---------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.23.1                         | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client |
| emr-5.23.0                         | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client |



| Amazon-EMR-Version<br>sbezeichnung | Sqoop-Version | Mit Sqoop installierte<br>Komponenten                                                                                                                                                                                                                                             |
|------------------------------------|---------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.22.0                         | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client |
| emr-5.21.2                         | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client |

| Amazon-EMR-Version<br>sbezeichnung | Sqoop-Version | Mit Sqoop installierte<br>Komponenten                                                                                                                                                                                                                                             |
|------------------------------------|---------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.21.1                         | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client |
| emr-5.21.0                         | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client |

| Amazon-EMR-Version<br>sbezeichnung | Sqoop-Version | Mit Sqoop installierte<br>Komponenten                                                                                                                                                                                                                                             |
|------------------------------------|---------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.20.1                         | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client |
| emr-5.20.0                         | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client |

| Amazon-EMR-Version<br>sbezeichnung | Sqoop-Version | Mit Sqoop installierte<br>Komponenten                                                                                                                                                                                                                                             |
|------------------------------------|---------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.19.1                         | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client |
| emr-5.19.0                         | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client |

| Amazon-EMR-Version<br>sbezeichnung | Sqoop-Version | Mit Sqoop installierte<br>Komponenten                                                                                                                                                                                                                                             |
|------------------------------------|---------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.18.1                         | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client |
| emr-5.18.0                         | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client |

| Amazon-EMR-Version<br>sbezeichnung | Sqoop-Version | Mit Sqoop installierte<br>Komponenten                                                                                                                                                                                                                                             |
|------------------------------------|---------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.17.2                         | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client |
| emr-5.17.1                         | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client |

| Amazon-EMR-Version<br>sbezeichnung | Sqoop-Version | Mit Sqoop installierte<br>Komponenten                                                                                                                                                                                                                                             |
|------------------------------------|---------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.17.0                         | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client |
| emr-5.16.1                         | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client |

| Amazon-EMR-Version<br>sbezeichnung | Sqoop-Version | Mit Sqoop installierte<br>Komponenten                                                                                                                                                                                                                                             |
|------------------------------------|---------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.16.0                         | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client |
| emr-5.15.1                         | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client |



| Amazon-EMR-Version<br>sbezeichnung | Sqoop-Version | Mit Sqoop installierte<br>Komponenten                                                                                                                                                                                                                                             |
|------------------------------------|---------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.15.0                         | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client |
| emr-5.14.2                         | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client |

| Amazon-EMR-Version<br>sbezeichnung | Sqoop-Version | Mit Sqoop installierte<br>Komponenten                                                                                                                                                                                                                                             |
|------------------------------------|---------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.14.1                         | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client |
| emr-5.14.0                         | 1.4.7         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client |

| Amazon-EMR-Version<br>sbezeichnung | Sqoop-Version | Mit Sqoop installierte<br>Komponenten                                                                                                                                                                                                                                             |
|------------------------------------|---------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.13.1                         | 1.4.6         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client |
| emr-5.13.0                         | 1.4.6         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client |

| Amazon-EMR-Version<br>sbezeichnung | Sqoop-Version | Mit Sqoop installierte<br>Komponenten                                                                                                                                                                                                                                             |
|------------------------------------|---------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.12.3                         | 1.4.6         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client |
| emr-5.12.2                         | 1.4.6         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client |

| Amazon-EMR-Version<br>sbezeichnung | Sqoop-Version | Mit Sqoop installierte<br>Komponenten                                                                                                                                                                                                                                                                               |
|------------------------------------|---------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.12.1                         | 1.4.6         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-ma-<br>pred, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>https-server, hadoop-kms-<br>server, hadoop-yarn-nodemanager,<br>hadoop-yarn-resourceman-<br>ager, hadoop-yarn-timeline-server,<br>mysql-server, sqoop-client    |
| emr-5.12.0                         | 1.4.6         | emrfs, emr-ddb, emr-goodies,<br>hadoop-client, hadoop-ma-<br>pred, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>https-server, hadoop-kms-<br>server, hadoop-yarn-nodemanager,<br>hadoop-yarn-resourceman-<br>ager, hadoop-yarn-timeline-server,<br>mysql-server, sqoop-client |

| Amazon-EMR-Version<br>sbezeichnung | Sqoop-Version | Mit Sqoop installierte<br>Komponenten                                                                                                                                                                                                                                             |
|------------------------------------|---------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.11.4                         | 1.4.6         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client |
| emr-5.11.3                         | 1.4.6         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client |

| Amazon-EMR-Version<br>sbezeichnung | Sqoop-Version | Mit Sqoop installierte<br>Komponenten                                                                                                                                                                                                                                             |
|------------------------------------|---------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.11.2                         | 1.4.6         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client |
| emr-5.11.1                         | 1.4.6         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client |

| Amazon-EMR-Version<br>sbezeichnung | Sqoop-Version | Mit Sqoop installierte<br>Komponenten                                                                                                                                                                                                                                              |
|------------------------------------|---------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.11.0                         | 1.4.6         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mahout, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client |
| emr-5.10.1                         | 1.4.6         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mahout, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client |



| Amazon-EMR-Version<br>sbezeichnung | Sqoop-Version | Mit Sqoop installierte<br>Komponenten                                                                                                                                                                                                                                             |
|------------------------------------|---------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.10.0                         | 1.4.6         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client |
| emr-5.9.1                          | 1.4.6         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client |

| Amazon-EMR-Version<br>sbezeichnung | Sqoop-Version | Mit Sqoop installierte<br>Komponenten                                                                                                                                                                                                                                             |
|------------------------------------|---------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.9.0                          | 1.4.6         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client |
| emr-5.8.3                          | 1.4.6         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client |

| Amazon-EMR-Version<br>sbezeichnung | Sqoop-Version | Mit Sqoop installierte<br>Komponenten                                                                                                                                                                                                                                             |
|------------------------------------|---------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.8.2                          | 1.4.6         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client |
| emr-5.8.1                          | 1.4.6         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client |

| Amazon-EMR-Version<br>sbezeichnung | Sqoop-Version | Mit Sqoop installierte<br>Komponenten                                                                                                                                                                                                                                             |
|------------------------------------|---------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.8.0                          | 1.4.6         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client |
| emr-5.7.1                          | 1.4.6         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client |

| Amazon-EMR-Version<br>sbezeichnung | Sqoop-Version | Mit Sqoop installierte<br>Komponenten                                                                                                                                                                                                                                             |
|------------------------------------|---------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.7.0                          | 1.4.6         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client |
| emr-5.6.1                          | 1.4.6         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client |

| Amazon-EMR-Version<br>sbezeichnung | Sqoop-Version | Mit Sqoop installierte<br>Komponenten                                                                                                                                                                                                                                             |
|------------------------------------|---------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.6.0                          | 1.4.6         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client |
| emr-5.5.4                          | 1.4.6         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client                              |
| emr-5.5.3                          | 1.4.6         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client                              |

| Amazon-EMR-Version<br>sbezeichnung | Sqoop-Version | Mit Sqoop installierte<br>Komponenten                                                                                                                                                                                                                |
|------------------------------------|---------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.5.2                          | 1.4.6         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client |
| emr-5.5.1                          | 1.4.6         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client |
| emr-5.5.0                          | 1.4.6         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client |

| Amazon-EMR-Version<br>sbezeichnung | Sqoop-Version | Mit Sqoop installierte<br>Komponenten                                                                                                                                                                                                                 |
|------------------------------------|---------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.4.1                          | 1.4.6         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client |
| emr-5.4.0                          | 1.4.6         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client |
| emr-5.3.2                          | 1.4.6         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client |



| Amazon-EMR-Version<br>sbezeichnung | Sqoop-Version | Mit Sqoop installierte<br>Komponenten                                                                                                                                                                                                                 |
|------------------------------------|---------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.3.1                          | 1.4.6         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client |
| emr-5.3.0                          | 1.4.6         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client |
| emr-5.2.3                          | 1.4.6         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client |

| Amazon-EMR-Version<br>sbezeichnung | Sqoop-Version | Mit Sqoop installierte<br>Komponenten                                                                                                                                                                                                                 |
|------------------------------------|---------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.2.2                          | 1.4.6         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client |
| emr-5.2.1                          | 1.4.6         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client |
| emr-5.2.0                          | 1.4.6         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client |

| Amazon-EMR-Version<br>sbezeichnung | Sqoop-Version | Mit Sqoop installierte<br>Komponenten                                                                                                                                                                                                                 |
|------------------------------------|---------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.1.1                          | 1.4.6         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client |
| emr-5.1.0                          | 1.4.6         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client |
| emr-5.0.3                          | 1.4.6         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client |

| Amazon-EMR-Version<br>sbezeichnung | Sqoop-Version | Mit Sqoop installierte<br>Komponenten                                                                                                                                                                                                                 |
|------------------------------------|---------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.0.2                          | 1.4.6         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client |
| emr-5.0.1                          | 1.4.6         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client |
| emr-5.0.0                          | 1.4.6         | emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client |

# TensorFlow

TensorFlow ist eine symbolische Open-Source-Mathematikbibliothek für maschinelle Intelligenz und Deep-Learning-Anwendungen. Weitere Informationen finden Sie [TensorFlow auf der Website](#). TensorFlow ist mit der Amazon EMR-Release-Version 5.17.0 und höher verfügbar.

In der folgenden Tabelle sind die Version von TensorFlow aufgeführt, die in der neuesten Version der Amazon EMR 7.x-Serie enthalten ist, zusammen mit den Komponenten, mit denen Amazon EMR installiert wird. TensorFlow

Informationen zur Version der Komponenten, mit denen TensorFlow in dieser Version installiert wurde, finden Sie unter Komponentenversionen von [Version 7.1.0](#).

TensorFlow Versionsinformationen für emr-7.1.0

| Amazon-EMR-Version<br>sbezeichnung | TensorFlow Version | Komponenten, die mit installiert wurden TensorFlow                                                                                                                                                                                         |
|------------------------------------|--------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-7.1.0                          | TensorFlow 2.11.0  | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow |

In der folgenden Tabelle sind die Version von TensorFlow aufgeführt, die in der neuesten Version der Amazon EMR 6.x-Serie enthalten ist, zusammen mit den Komponenten, mit denen Amazon EMR installiert wird. TensorFlow

Informationen zur Version der Komponenten, mit denen TensorFlow in dieser Version installiert wurde, finden Sie unter Komponentenversionen von [Version 6.15.0](#).

## TensorFlow Versionsinformationen für emr-6.15.0

| Amazon-EMR-Version<br>sbezeichnung | TensorFlow Version | Komponenten, die mit installiert wurden TensorFlow                                                                                                                                                                                          |
|------------------------------------|--------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.15.0                         | TensorFlow 2.11.0  | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow |

In der folgenden Tabelle sind die Version von TensorFlow aufgeführt, die in der neuesten Version der Amazon EMR 5.x-Serie enthalten ist, zusammen mit den Komponenten, mit denen Amazon EMR installiert wird. TensorFlow

Informationen zur Version der Komponenten, mit denen TensorFlow in dieser Version installiert wurde, finden Sie unter [Komponentenversionen von Version 5.36.2](#).

## TensorFlow Versionsinformationen für emr-5.36.2

| Amazon-EMR-Version<br>sbezeichnung | TensorFlow Version | Komponenten, die mit installiert wurden TensorFlow                                                                                                                                                                                          |
|------------------------------------|--------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.36.2                         | TensorFlow 2.4.1   | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow |

# TensorFlow Builds nach Amazon EC2 EC2-Instanztyp

Amazon EMR verwendet je nach den Instance-Typen, die Sie für Ihren Cluster auswählen, unterschiedliche Builds der TensorFlow Bibliothek. In der folgenden Tabelle sind die Builds nach Instance-Typ gruppiert aufgelistet.

| EC2-Instance-Typen | TensorFlow erstellen                                                                                                                                                                                                                                                                                      |
|--------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| M5 und C5          | TensorFlow 1.9.0 mit Intel MKL-Optimierung                                                                                                                                                                                                                                                                |
| P2                 | TensorFlow 1.9.0 mit CUDA 9.2, cuDNN 7.1                                                                                                                                                                                                                                                                  |
| P3                 | TensorFlow 1.9.0 mit CUDA 9.2, cuDNN 7.1, NCCL 2.2.13<br><br><a href="#">Nvidia NCCL</a> ist nur auf P3-Instances verfügbar . Endbenutzer-Lizenzvereinbarung (EULA): Durch die Verwendung von Nvidia-Komponenten auf Amazon EMR, stimmen Sie den Bedingungen der <a href="#">EULA für das Produkt</a> zu. |
| Alle anderen       | TensorFlow 1.9.0                                                                                                                                                                                                                                                                                          |

## Sicherheit

Zusätzlich zu den Anweisungen [unter TensorFlow Sicheres Verwenden](#) empfehlen wir, Ihren Cluster in einem privaten Subnetz zu starten, um den Zugriff auf vertrauenswürdige Quellen zu beschränken. Weitere Informationen über [Amazon-VPC-Optionen](#) finden Sie im Verwaltungshandbuch für Amazon EMR.

## Verwenden TensorBoard

TensorBoard ist eine Suite von Visualisierungstools für TensorFlow Programme. Weitere Informationen finden Sie unter [TensorBoard: Visualisiertes Lernen](#) auf der Tensorflow-Website.

Für die Verwendung TensorBoard mit Amazon EMR müssen Sie TensorBoard auf dem Cluster-Masterknoten beginnen.

## So verwenden Sie TensorBoard mit TensorFlow auf Amazon EMR

1. Stellen Sie mit dem Master-Knoten Ihres Clusters mithilfe von SSH eine Verbindung her. Weitere Informationen finden Sie unter [Mit SSH eine Verbindung zum Hauptknoten herstellen](#) im Verwaltungshandbuch für Amazon EMR.
2. Geben Sie den folgenden Befehl ein, um TensorBoard auf dem Master-Knoten zu starten. Ersetzen Sie `/my/log/directory` mit dem Verzeichnis auf dem Master-Knoten, in dem Sie unter Verwendung einer Summary Writer-Operation die Übersichtsdaten gespeichert haben.

### Amazon EMR 5.19.0 and later

```
python3 -m tensorboard.main --logdir=/home/hadoop/tensor --bind_all
```

### Amazon EMR 5.18.1 and earlier

```
python3 -m tensorboard.main --logdir=/my/log/dir
```

Standardmäßig verwendet der Master-Knoten Port 6006 und den öffentlichen DNS-Namen des Masters. Nach dem Start TensorBoard wird in der Befehlszeilenausgabe die URL angezeigt, mit der eine Verbindung hergestellt werden kann TensorBoard, wie im folgenden Beispiel gezeigt:

```
TensorBoard 1.9.0 at http://master-public-dns-name:6006 (Press CTRL+C to quit)
```

3. Richten Sie den Zugriff auf Webschnittstellen auf dem Master-Knoten von vertrauenswürdigen Clients aus ein. Weitere Informationen finden Sie unter [Anzeigen von auf Amazon-EMR-Clustern gehosteten Webschnittstellen](#) im Verwaltungshandbuch für Amazon EMR.
4. Öffnet TensorBoard unter `http://master-public-dns-name:6006`.

## TensorFlow Verlauf der Veröffentlichung

In der folgenden Tabelle sind die Versionen von TensorFlow aufgeführt, die in jeder Release-Version von Amazon EMR enthalten sind, zusammen mit den Komponenten, die mit der Anwendung installiert wurden. Informationen zu den Komponentenversionen in den einzelnen Versionen finden Sie im Abschnitt [Komponentenversion für Ihre Version in Amazon-EMR-7.x-Versionen](#), [Amazon-EMR-6.x-Versionen](#) oder [Amazon-EMR-5.x-Versionen](#).



## TensorFlow Versionsinformationen

| Amazon-EMR-Version<br>sbezeichnung | TensorFlow Version | Komponenten, die installiert<br>wurden mit TensorFlow                                                                                                                                                                                     |
|------------------------------------|--------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.36.2                         | 2.4.1              | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow |
| emr-7.1.0                          | 2.11.0             | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow |
| emr-7.0.0                          | 2.11.0             | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow |
| emr-6.15.0                         | 2.11.0             | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-                                                                                                                                                     |

| Amazon-EMR-Version<br>sbezeichnung | TensorFlow Version | Komponenten, die installiert<br>wurden mit TensorFlow                                                                                                                                                                                                                       |
|------------------------------------|--------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                                    |                    | hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema-<br>nager, hadoop-yarn-resour-<br>cemanager, hadoop-yarn-<br>timeline-server, tensorflow                                                                                                 |
| emr-6.14.0                         | 2.11.0             | emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema-<br>nager, hadoop-yarn-resour-<br>cemanager, hadoop-yarn-<br>timeline-server, tensorflow |
| emr-6.13.0                         | 2.11.0             | emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema-<br>nager, hadoop-yarn-resour-<br>cemanager, hadoop-yarn-<br>timeline-server, tensorflow |

| Amazon-EMR-Version<br>sbezeichnung | TensorFlow Version | Komponenten, die installiert<br>wurden mit TensorFlow                                                                                                                                                                                       |
|------------------------------------|--------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.12.0                         | 2.11.0             | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow |
| emr-6.11.1                         | 2.11.0             | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow |
| emr-6.11.0                         | 2.11.0             | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow |

| Amazon-EMR-Version<br>sbezeichnung | TensorFlow Version | Komponenten, die installiert<br>wurden mit TensorFlow                                                                                                                                                                                      |
|------------------------------------|--------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.10.1                         | 2.11.0             | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow |
| emr-6.10.0                         | 2.11.0             | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow |
| emr-6.9.1                          | 2.10.0             | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow |

| Amazon-EMR-Version<br>sbezeichnung | TensorFlow Version | Komponenten, die installiert<br>wurden mit TensorFlow                                                                                                                                                                                       |
|------------------------------------|--------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.9.0                          | 2.10.0             | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow |
| emr-6.8.1                          | 2.9.1              | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow |
| emr-6.8.0                          | 2.9.1              | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow |

| Amazon-EMR-Version<br>sbezeichnung | TensorFlow Version | Komponenten, die installiert<br>wurden mit TensorFlow                                                                                                                                                                                       |
|------------------------------------|--------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.7.0                          | 2.4.1              | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow |
| emr-5.36.1                         | 2.4.1              | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow |
| emr-5.36.0                         | 2.4.1              | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow |

| Amazon-EMR-Version<br>sbezeichnung | TensorFlow Version | Komponenten, die installiert<br>wurden mit TensorFlow                                                                                                                                                                                       |
|------------------------------------|--------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.6.0                          | 2.4.1              | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow |
| emr-5.35.0                         | 2.4.1              | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow |
| emr-6.5.0                          | 2.4.1              | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow |

| Amazon-EMR-Version<br>sbezeichnung | TensorFlow Version | Komponenten, die installiert<br>wurden mit TensorFlow                                                                                                                                                                                      |
|------------------------------------|--------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.4.0                          | 2.4.1              | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow |
| emr-6.3.1                          | 2.4.1              | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow |
| emr-6.3.0                          | 2.4.1              | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow |



| Amazon-EMR-Version<br>sbezeichnung | TensorFlow Version | Komponenten, die installiert<br>wurden mit TensorFlow                                                                                                                                                                                      |
|------------------------------------|--------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.2.1                          | 2.3.1              | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow |
| emr-6.2.0                          | 2.3.1              | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow |
| emr-6.1.1                          | 2.1.0              | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow |

| Amazon-EMR-Version<br>sbezeichnung | TensorFlow Version | Komponenten, die installiert<br>wurden mit TensorFlow                                                                                                                                                                                      |
|------------------------------------|--------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.1.0                          | 2.1.0              | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow |
| emr-6.0.1                          | 1.14.0             | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow |
| emr-6.0.0                          | 1.14.0             | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow |

| Amazon-EMR-Version<br>sbezeichnung | TensorFlow Version | Komponenten, die installiert<br>wurden mit TensorFlow                                                                                                                                                                                       |
|------------------------------------|--------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.34.0                         | 2.4.1              | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow |
| emr-5.33.1                         | 2.4.1              | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow |
| emr-5.33.0                         | 2.4.1              | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow |

| Amazon-EMR-Version<br>sbezeichnung | TensorFlow Version | Komponenten, die installiert<br>wurden mit TensorFlow                                                                                                                                                                                      |
|------------------------------------|--------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.32.1                         | 2.3.1              | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow |
| emr-5.32.0                         | 2.3.1              | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow |
| emr-5.31.1                         | 2.1.0              | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow |

| Amazon-EMR-Version<br>sbezeichnung | TensorFlow Version | Komponenten, die installiert<br>wurden mit TensorFlow                                                                                                                                                                                      |
|------------------------------------|--------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.31.0                         | 2.1.0              | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow |
| emr-5.30.2                         | 1.14.0             | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow |
| emr-5.30.1                         | 1.14.0             | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow |

| Amazon-EMR-Version<br>sbezeichnung | TensorFlow Version | Komponenten, die installiert<br>wurden mit TensorFlow                                                                                                                                                                                      |
|------------------------------------|--------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.30.0                         | 1.14.0             | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow |
| emr-5.29.0                         | 1.14.0             | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow |
| emr-5.28.1                         | 1.14.0             | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow |

| Amazon-EMR-Version<br>sbezeichnung | TensorFlow Version | Komponenten, die installiert<br>wurden mit TensorFlow                                                                                                                                                                                       |
|------------------------------------|--------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.28.0                         | 1.14.0             | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow |
| emr-5.27.1                         | 1.14.0             | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow |
| emr-5.27.0                         | 1.14.0             | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow |

| Amazon-EMR-Version<br>sbezeichnung | TensorFlow Version | Komponenten, die installiert<br>wurden mit TensorFlow                                                                                                                                                                                      |
|------------------------------------|--------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.26.0                         | 1.13.1             | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow |
| emr-5.25.0                         | 1.13.1             | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow |
| emr-5.24.1                         | 1.12.0             | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow |



| Amazon-EMR-Version<br>sbezeichnung | TensorFlow Version | Komponenten, die installiert<br>wurden mit TensorFlow                                                                                                                                                                                       |
|------------------------------------|--------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.24.0                         | 1.12.0             | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow |
| emr-5.23.1                         | 1.12.0             | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow |
| emr-5.23.0                         | 1.12.0             | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow |

| Amazon-EMR-Version<br>sbezeichnung | TensorFlow Version | Komponenten, die installiert<br>wurden mit TensorFlow                                                                                                                                                                                      |
|------------------------------------|--------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.22.0                         | 1.12.0             | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow |
| emr-5.21.2                         | 1.12.0             | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow |
| emr-5.21.1                         | 1.12.0             | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow |

| Amazon-EMR-Version<br>sbezeichnung | TensorFlow Version | Komponenten, die installiert<br>wurden mit TensorFlow                                                                                                                                                                                      |
|------------------------------------|--------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.21.0                         | 1.12.0             | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow |
| emr-5.20.1                         | 1.12.0             | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow |
| emr-5.20.0                         | 1.12.0             | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow |

| Amazon-EMR-Version<br>sbezeichnung | TensorFlow Version | Komponenten, die installiert<br>wurden mit TensorFlow                                                                                                                                                                                      |
|------------------------------------|--------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.19.1                         | 1.11.0             | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow |
| emr-5.19.0                         | 1.11.0             | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow |
| emr-5.18.1                         | 1.9.0              | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow |

| Amazon-EMR-Version<br>sbezeichnung | TensorFlow Version | Komponenten, die installiert<br>wurden mit TensorFlow                                                                                                                                                                                      |
|------------------------------------|--------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.18.0                         | 1.9.0              | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow |
| emr-5.17.2                         | 1.9.0              | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow |
| emr-5.17.1                         | 1.9.0              | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow |

| Amazon-EMR-Version<br>sbezeichnung | TensorFlow Version | Komponenten, die installiert<br>wurden mit TensorFlow                                                                                                                                                                                      |
|------------------------------------|--------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.17.0                         | 1.9.0              | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow |

# Apache Tez

Apache Tez ist ein Framework zum Erstellen eines komplexen azyklisch gerichteten Diagramms (DAG) von Aufgaben für die Verarbeitung von Daten ermöglicht. In einigen Anwendungsfällen können Sie es als Alternative zu Hadoop MapReduce verwenden. Sie können beispielsweise Pig- und Hive-Workflows mit Hadoop ausführen MapReduce oder Tez als Ausführungs-Engine verwenden. Weitere Informationen finden Sie unter <https://tez.apache.org/>. Amazon-EMR-Versionen 4.7.0 und höher enthalten Tez.

Die folgende Tabelle listet die Version von Tez auf, die in der neuesten Version der Amazon-EMR-Serie 7.x enthalten ist, zusammen mit den Komponenten, die Amazon EMR mit Tez installiert.

[Die Version der Komponenten, die in dieser Version mit Tez installiert wurden, finden Sie unter Komponentenversionen von Version 7.1.0.](#)

Tez-Versionsinformationen für emr-7.1.0

| Amazon-EMR-Version<br>sbezeichnung | Tez-Version | Mit Tez installierte Komponent<br>en                                                                                                                                                                                                                                                   |
|------------------------------------|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-7.1.0                          | Tez 0.10.2  | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn,<br>tez-on-worker |

Die folgende Tabelle listet die Version von Tez auf, die in der neuesten Version der Amazon-EMR-6.x-Serie enthalten ist, zusammen mit den Komponenten, die Amazon EMR mit Tez installiert.

Die Version der Komponenten, die mit Tez in dieser Version installiert wurden, finden Sie unter [Komponentenversionen der Version 6.15.0.](#)

## Tez-Versionsinformationen für emr-6.15.0

| Amazon-EMR-Version<br>sbezeichnung | Tez-Version | Mit Tez installierte Komponent<br>en                                                                                                                                                                                                                                                   |
|------------------------------------|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.15.0                         | Tez 0.10.2  | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn,<br>tez-on-worker |

Die folgende Tabelle listet die Version von Tez auf, die in der neuesten Version der Amazon-EMR-5.x-Serie enthalten ist, zusammen mit den Komponenten, die Amazon EMR mit Tez installiert.

[Informationen zur Version der Komponenten, die in dieser Version mit Tez installiert wurden, finden Sie unter Komponentenversionen von Version 5.36.2.](#)

## Tez-Versionsinformationen für emr-5.36.2

| Amazon-EMR-Version<br>sbezeichnung | Tez-Version | Mit Tez installierte Komponent<br>en                                                                                                                                                                                                                                 |
|------------------------------------|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.36.2                         | Tez 0.9.2   | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |



## Themen

- [Erstellen eines Clusters mit Tez](#)
- [Konfigurieren von Tez](#)
- [Tez-Webbenutzeroberfläche](#)
- [Timeline-Server](#)
- [Tez-Versionsverlauf](#)

## Erstellen eines Clusters mit Tez

Um Tez zu installieren, wählen Sie Apache Tez als Anwendung, wenn Sie Ihren Cluster erstellen.

So erstellen Sie einen Cluster mit Tez über die Konsole

1. Navigieren Sie zur neuen Amazon-EMR-Konsole und wählen Sie in der Seitennavigation die Option [Zur alten Konsole wechseln](#) aus. Weitere Informationen darüber, was Sie erwartet, wenn Sie zur alten Konsole wechseln, finden Sie unter [Verwenden der alten Konsole](#).
2. Wählen Sie **Create Cluster** (Cluster erstellen) und **Go to advanced options** (Zu erweiterten Optionen) aus.
3. Wählen Sie unter **Softwarekonfiguration** das Release `emr-4.7.0` oder höher aus.
4. Wählen Sie **Tez** zusammen mit anderen Anwendungen aus, die Amazon EMR installieren soll.
5. Wählen Sie nach Bedarf weitere Optionen und anschließend **Create cluster** (Cluster erstellen) aus.

Um einen Cluster mit Tez zu erstellen, verwenden Sie AWS CLI

- Verwenden Sie den `create-cluster`-Befehl zusammen mit der `-- applications`-Option, um Tez festzulegen. Im folgenden Beispiel wird ein Cluster mit installiertem Tez erstellt.

### Note

Linux-Zeilenfortsetzungszeichen (`\`) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (`^`).

```
aws emr create-cluster --name "Cluster with Tez" --release-label emr-7.1.0 \  
--applications Name=Tez --ec2-attributes KeyName=myKey \  
--instance-type m5.xlarge --instance-count 3 --use-default-roles
```

## Konfigurieren von Tez

Sie können Tez anpassen, indem Sie mit der Konfigurationsklassifizierung `tez-site` Einstellungen in der Konfigurationsdatei `tez-site.xml` konfigurieren. Weitere Informationen finden Sie [TezConfiguration](#) in der Apache Tez-Dokumentation. Verwenden Sie entsprechend die `hive-site`- und `pig-properties`-Konfigurationsklassifizierungen, um zu ändern, dass Hive oder Pig die Tez-Ausführungs-Engine verwendet. Beispiele werden unten angezeigt.

## Beispielkonfiguration

Example Beispiel: Anpassen der Tez-Stammprotokollierungsebene und Festlegen von Tez als Ausführungs-Engine für Hive und Pig

Mit dem unten angezeigten `create-cluster`-Beispielbefehl wird ein Cluster mit installierten Tez, Hive und Pig erstellt. Der Befehl verweist auf eine Datei die in Amazon S3 `myConfig.json` gespeichert ist und die Eigenschaften für die `tez-site`-Klassifizierung angibt, die `tez.am.log.level` auf `DEBUG` setzt und die Ausführungs-Engine für Hive und Pig mit den Konfigurationsklassifizierungen `hive-site` und `pig-properties` auf Tez festlegt.

### Note

Linux-Zeilenfortsetzungszeichen (`\`) sind aus Gründen der Lesbarkeit enthalten. Sie können entfernt oder in Linux-Befehlen verwendet werden. Entfernen Sie sie unter Windows oder ersetzen Sie sie durch ein Caret-Zeichen (`^`).

```
aws emr create-cluster --release-label emr-7.1.0 \  
--applications Name=Tez Name=Hive Name=Pig --ec2-attributes KeyName=myKey \  
--instance-type m5.xlarge --instance-count 3 \  
--configurations https://s3.amazonaws.com/mybucket/myfolder/myConfig.json --use-  
default-roles
```

Beispielinhalte von `myConfig.json` werden unten angezeigt.

```
[
  {
    "Classification": "tez-site",
    "Properties": {
      "tez.am.log.level": "DEBUG"
    }
  },
  {
    "Classification": "hive-site",
    "Properties": {
      "hive.execution.engine": "tez"
    }
  },
  {
    "Classification": "pig-properties",
    "Properties": {
      "exectype": "tez"
    }
  }
]
```

#### Note

Ab Amazon-EMR-Version 5.21.0 können Sie Cluster-Konfigurationen überschreiben und zusätzliche Konfigurationsklassifikationen für jede Instance-Gruppe in einem ausgeführten Cluster angeben. Dazu verwenden Sie die Amazon EMR-Konsole, das AWS Command Line Interface (AWS CLI) oder das AWS SDK. Weitere Informationen finden Sie unter [Angeben einer Konfiguration für eine Instance-Gruppe in einem aktiven Cluster](#).

## Asynchrone Öffnung von Tez-Splits

Wenn der Tabellenpfad eine große Anzahl kleiner Dateien enthält und eine Abfrage versucht, sie alle zu lesen, wird jede kleine Datei, die jeder einzelnen Aufteilung entspricht, zu einem gruppierten Tez-Split zusammengefasst. Ein einzelner Mapper verarbeitet dann den einzelnen gruppierten Tez-Split. Da die Ausführung synchron ist, wird jeder einzelne Split im Rahmen des gruppierten Splits einzeln verarbeitet. Dafür ist erforderlich, dass `RecordReader`-Objekte die Splits synchron verarbeiten.

| Name                                                    | Klassifizierung       | Beschreibung                                                                                                                                                                                                                                                                                                                                                                                                  |
|---------------------------------------------------------|-----------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <code>tez.grouping.split<br/>.init.threads</code>       | <code>tez-site</code> | Gibt die Anzahl der Daemon-Threads an, die Tez verwendet, um die <code>RecordReaders</code> vorab zu initiieren und Splits zu öffnen. Für ACID-Tabellen ist der unterstützte Höchstwert von <code>tez.grouping.split.init.threads</code> 1.                                                                                                                                                                   |
| <code>tez.grouping.split<br/>.init.recordreaders</code> | <code>tez-site</code> | Gibt die Anzahl an <code>RecordReaders</code> an, die von den Daemon-Threads vorinitialisiert werden sollen. Dies kann hilfreich sein, wenn ein gruppierter Tez-Split eine große Anzahl von <code>InputSplits</code> enthält. Die Initialisierung von <code>RecordReaders</code> zur Verarbeitung dieser Eingabe-Splits kann asynchron mit Daemon-Threads anstelle einer sequentiellen Verarbeitung erfolgen. |

## Benchmarking für das asynchrone Öffnen von Tez-Splits

Wir haben die folgenden Umgebungen und Konfigurationen für das Benchmarking der asynchronen Split-Öffnungsfunktion von Tez verwendet:

- Benchmark-Umgebung – Amazon-EMR-Cluster mit einem Primärknoten, der `m5.16xlarge` verwendet, und 16 Core-Knoten, die `m5.16xlarge` verwenden.

- Benchmark-Konfigurationen – Um das Szenario für das Benchmarking zu simulieren, bei dem sich eine große Anzahl von Eingabe-Splits in einem einzigen gruppierten Tez-Split befinden, ist `tez.grouping.split-count` auf 1 eingestellt.
- Für das Benchmarking verwendete Tabelle – Die Tabelle enthält 200 Partitionen, wobei jede Partition eine einzelne Datei enthält. Der Benchmark wird durchgeführt, wenn diese Tabelle CSV-Dateien enthält und wenn diese Tabelle Parquet-Dateien enthält. Hive-Abfrage für das Benchmarking: zehnmal `SELECT COUNT(*)` aus der Tabelle, und es wird die durchschnittliche Laufzeit berechnet.
- Konfigurationen zur Aktivierung des asynchronen geteilten Öffnens von Tez – wie folgt:
  - `tez.grouping.split.init.threads = 4`
  - `tez.grouping.split.init.recordreaders = 10`

| Datensatz         | Feature deaktiviert<br>(Basisversion) | Feature aktiviert | Verbesserung |
|-------------------|---------------------------------------|-------------------|--------------|
| CSV-Datensatz     | 90,26 Sekunden                        | 79,20 Sekunden    | 12,25%       |
| Parquet-Datensatz | 54,67 Sekunden                        | 42,23 Sekunden    | 22,75%       |

## Tez-Webbenutzeroberfläche

Tez verfügt über eine eigene Webbenutzeroberfläche. Die Webbenutzeroberfläche finden Sie unter der folgenden URL.

```
http://masterDNS:8080/tez-ui
```

Um die Registerkarte Hive-Abfragen auf der Tez-Weboberfläche zu aktivieren, stellen Sie die folgende Konfiguration ein.

```
[
  {
    "Classification": "hive-site",
    "Properties": {
      "hive.exec.pre.hooks": "org.apache.hadoop.hive.q1.hooks.ATSHook",
      "hive.exec.post.hooks": "org.apache.hadoop.hive.q1.hooks.ATSHook",
      "hive.exec.failure.hooks": "org.apache.hadoop.hive.q1.hooks.ATSHook"
    }
  }
]
```

```

    }
  }
]
```

Sie können die Details der Tez-, Spark- und YARN-Anwendungsbenutzeroberflächen auch mithilfe von Links auf der Registerkarte Anwendungsbenutzeroberflächen der Detailseite eines Clusters in der Konsole anzeigen. Die Benutzeroberflächen (UI) von Amazon EMR werden außerhalb des Clusters gehostet und sind verfügbar, nachdem der Cluster beendet wurde. Sie müssen keine SSH-Verbindung oder Web-Proxys einrichten, was Ihnen die Fehlerbehebung und Analyse aktiver Aufgaben und des Aufgabenverlaufs erleichtert.

Weitere Informationen finden Sie unter [Anwendungsverlauf anzeigen](#) im Verwaltungshandbuch für Amazon EMR.

## Timeline-Server

Der YARN Timeline Server ist so konfiguriert, dass er ausgeführt wird, wenn Tez installiert ist. Um Jobs anzusehen, die über Tez oder MapReduce Execution Engines mit dem Timeline Server eingereicht wurden, rufen Sie die Weboberfläche mithilfe der URL auf. `http://master-public-DNS:8188` Weitere Informationen finden Sie unter [Anzeigen von auf Amazon-EMR-Clustern gehosteten Webschnittstellen](#) im Verwaltungshandbuch für Amazon EMR.

## Tez-Versionsverlauf

In der folgenden Tabelle sind die Version von Tez aufgeführt, die in jeder Release-Version von Amazon EMR enthalten ist, zusammen mit den Komponenten, die mit der Anwendung installiert wurden. Informationen zu den Komponentenversionen in den einzelnen Versionen finden Sie im Abschnitt [Komponentenversion für Ihre Version in Amazon-EMR-7.x-Versionen](#), [Amazon-EMR-6.x-Versionen](#) oder [Amazon-EMR-5.x-Versionen](#).

Die Tez-Versionsinformationen

| Amazon-EMR-Version<br>sbezeichnung | Tez-Version | Mit Tez installierte Komponent<br>en                                                    |
|------------------------------------|-------------|-----------------------------------------------------------------------------------------|
| emr-5.36.2                         | 0.9.2       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop- |

| Amazon-EMR-Version<br>sbezeichnung | Tez-Version | Mit Tez installierte Komponent<br>en                                                                                                                                                                                                                |
|------------------------------------|-------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                                    |             | hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn                                                                                               |
| emr-7.1.0                          | 0.10.2      | emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn, tez-on-worker |
| emr-7.0.0                          | 0.10.2      | emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn, tez-on-worker |

| Amazon-EMR-Version<br>sbezeichnung | Tez-Version | Mit Tez installierte Komponent<br>en                                                                                                                                                                                                                                                   |
|------------------------------------|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.15.0                         | 0.10.2      | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn,<br>tez-on-worker |
| emr-6.14.0                         | 0.10.2      | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn,<br>tez-on-worker |
| emr-6.13.0                         | 0.10.2      | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn,<br>tez-on-worker |



| Amazon-EMR-Version<br>sbezeichnung | Tez-Version | Mit Tez installierte Komponent<br>en                                                                                                                                                                                                                                                   |
|------------------------------------|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.12.0                         | 0.10.2      | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn,<br>tez-on-worker |
| emr-6.11.1                         | 0.10.2      | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn,<br>tez-on-worker |
| emr-6.11.0                         | 0.10.2      | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn,<br>tez-on-worker |

| Amazon-EMR-Version<br>sbezeichnung | Tez-Version | Mit Tez installierte Komponent<br>en                                                                                                                                                                                                                                                   |
|------------------------------------|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.10.1                         | 0.10.2      | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn,<br>tez-on-worker |
| emr-6.10.0                         | 0.10.2      | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn,<br>tez-on-worker |
| emr-6.9.1                          | 0.10.2      | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn                   |

| Amazon-EMR-Version<br>sbezeichnung | Tez-Version | Mit Tez installierte Komponent<br>en                                                                                                                                                                                                                                 |
|------------------------------------|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.9.0                          | 0.10.2      | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-6.8.1                          | 0.9.2       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-6.8.0                          | 0.9.2       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |

| Amazon-EMR-Version<br>sbezeichnung | Tez-Version | Mit Tez installierte Komponent<br>en                                                                                                                                                                                                                                 |
|------------------------------------|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.7.0                          | 0.9.2       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-5.36.1                         | 0.9.2       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-5.36.0                         | 0.9.2       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |

| Amazon-EMR-Version<br>sbezeichnung | Tez-Version | Mit Tez installierte Komponent<br>en                                                                                                                                                                                                                                 |
|------------------------------------|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.6.0                          | 0.9.2       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-5.35.0                         | 0.9.2       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-6.5.0                          | 0.9.2       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |

| Amazon-EMR-Version<br>sbezeichnung | Tez-Version | Mit Tez installierte Komponent<br>en                                                                                                                                                                                                                                 |
|------------------------------------|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.4.0                          | 0.9.2       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-6.3.1                          | 0.9.2       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-6.3.0                          | 0.9.2       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |

| Amazon-EMR-Version<br>sbezeichnung | Tez-Version | Mit Tez installierte Komponent<br>en                                                                                                                                                                                                                                 |
|------------------------------------|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.2.1                          | 0.9.2       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-6.2.0                          | 0.9.2       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-6.1.1                          | 0.9.2       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |

| Amazon-EMR-Version<br>sbezeichnung | Tez-Version | Mit Tez installierte Komponent<br>en                                                                                                                                                                                                                                 |
|------------------------------------|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.1.0                          | 0.9.2       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-6.0.1                          | 0.9.2       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-6.0.0                          | 0.9.2       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |



| Amazon-EMR-Version<br>sbezeichnung | Tez-Version | Mit Tez installierte Komponent<br>en                                                                                                                                                                                                                                 |
|------------------------------------|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.34.0                         | 0.9.2       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-5.33.1                         | 0.9.2       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-5.33.0                         | 0.9.2       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |

| Amazon-EMR-Version<br>sbezeichnung | Tez-Version | Mit Tez installierte Komponent<br>en                                                                                                                                                                                                                                 |
|------------------------------------|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.32.1                         | 0.9.2       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-5.32.0                         | 0.9.2       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-5.31.1                         | 0.9.2       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |

| Amazon-EMR-Version<br>sbezeichnung | Tez-Version | Mit Tez installierte Komponent<br>en                                                                                                                                                                                                                                 |
|------------------------------------|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.31.0                         | 0.9.2       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-5.30.2                         | 0.9.2       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-5.30.1                         | 0.9.2       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |

| Amazon-EMR-Version<br>sbezeichnung | Tez-Version | Mit Tez installierte Komponent<br>en                                                                                                                                                                                                                                 |
|------------------------------------|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.30.0                         | 0.9.2       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-5.29.0                         | 0.9.2       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-5.28.1                         | 0.9.2       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |

| Amazon-EMR-Version<br>sbezeichnung | Tez-Version | Mit Tez installierte Komponent<br>en                                                                                                                                                                                                                                 |
|------------------------------------|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.28.0                         | 0.9.2       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-5.27.1                         | 0.9.2       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-5.27.0                         | 0.9.2       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |

| Amazon-EMR-Version<br>sbezeichnung | Tez-Version | Mit Tez installierte Komponent<br>en                                                                                                                                                                                                                                 |
|------------------------------------|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.26.0                         | 0.9.2       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-5.25.0                         | 0.9.2       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-5.24.1                         | 0.9.1       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |

| Amazon-EMR-Version<br>sbezeichnung | Tez-Version | Mit Tez installierte Komponent<br>en                                                                                                                                                                                                                                 |
|------------------------------------|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.24.0                         | 0.9.1       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-5.23.1                         | 0.9.1       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-5.23.0                         | 0.9.1       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |

| Amazon-EMR-Version<br>sbezeichnung | Tez-Version | Mit Tez installierte Komponent<br>en                                                                                                                                                                                                                                 |
|------------------------------------|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.22.0                         | 0.9.1       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-5.21.2                         | 0.9.1       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-5.21.1                         | 0.9.1       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |



| Amazon-EMR-Version<br>sbezeichnung | Tez-Version | Mit Tez installierte Komponent<br>en                                                                                                                                                                                                                                 |
|------------------------------------|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.21.0                         | 0.9.1       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-5.20.1                         | 0.9.1       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-5.20.0                         | 0.9.1       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |

| Amazon-EMR-Version<br>sbezeichnung | Tez-Version | Mit Tez installierte Komponent<br>en                                                                                                                                                                                                                                 |
|------------------------------------|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.19.1                         | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-5.19.0                         | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-5.18.1                         | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |

| Amazon-EMR-Version<br>sbezeichnung | Tez-Version | Mit Tez installierte Komponent<br>en                                                                                                                                                                                                                                 |
|------------------------------------|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.18.0                         | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-5.17.2                         | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-5.17.1                         | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |

| Amazon-EMR-Version<br>sbezeichnung | Tez-Version | Mit Tez installierte Komponent<br>en                                                                                                                                                                                                                                 |
|------------------------------------|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.17.0                         | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-5.16.1                         | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-5.16.0                         | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |

| Amazon-EMR-Version<br>sbezeichnung | Tez-Version | Mit Tez installierte Komponent<br>en                                                                                                                                                                                                                                 |
|------------------------------------|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.15.1                         | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-5.15.0                         | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-5.14.2                         | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |

| Amazon-EMR-Version<br>sbezeichnung | Tez-Version | Mit Tez installierte Komponent<br>en                                                                                                                                                                                                                                 |
|------------------------------------|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.14.1                         | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-5.14.0                         | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-5.13.1                         | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |

| Amazon-EMR-Version<br>sbezeichnung | Tez-Version | Mit Tez installierte Komponent<br>en                                                                                                                                                                                                                                 |
|------------------------------------|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.13.0                         | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-5.12.3                         | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-5.12.2                         | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |

| Amazon-EMR-Version<br>sbezeichnung | Tez-Version | Mit Tez installierte Komponent<br>en                                                                                                                                                                                                                                 |
|------------------------------------|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.12.1                         | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-5.12.0                         | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-5.11.4                         | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |



| Amazon-EMR-Version<br>sbezeichnung | Tez-Version | Mit Tez installierte Komponent<br>en                                                                                                                                                                                                                                 |
|------------------------------------|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.11.3                         | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-5.11.2                         | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-5.11.1                         | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |

| Amazon-EMR-Version<br>sbezeichnung | Tez-Version | Mit Tez installierte Komponent<br>en                                                                                                                                                                                                                                 |
|------------------------------------|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.11.0                         | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-5.10.1                         | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-5.10.0                         | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |

| Amazon-EMR-Version<br>sbezeichnung | Tez-Version | Mit Tez installierte Komponent<br>en                                                                                                                                                                                                                                 |
|------------------------------------|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.9.1                          | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-5.9.0                          | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-5.8.3                          | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |

| Amazon-EMR-Version<br>sbezeichnung | Tez-Version | Mit Tez installierte Komponent<br>en                                                                                                                                                                                                                                 |
|------------------------------------|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.8.2                          | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-5.8.1                          | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-5.8.0                          | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |

| Amazon-EMR-Version<br>sbezeichnung | Tez-Version | Mit Tez installierte Komponent<br>en                                                                                                                                                                                                                                 |
|------------------------------------|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.7.1                          | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-5.7.0                          | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-5.6.1                          | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |

| Amazon-EMR-Version<br>sbezeichnung | Tez-Version | Mit Tez installierte Komponent<br>en                                                                                                                                                                                                                                 |
|------------------------------------|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.6.0                          | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-5.5.4                          | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-5.5.3                          | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |

| Amazon-EMR-Version<br>sbezeichnung | Tez-Version | Mit Tez installierte Komponent<br>en                                                                                                                                                                                                                                 |
|------------------------------------|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.5.2                          | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-5.5.1                          | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-5.5.0                          | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |

| Amazon-EMR-Version<br>sbezeichnung | Tez-Version | Mit Tez installierte Komponent<br>en                                                                                                                                                                                                                                 |
|------------------------------------|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.4.1                          | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-5.4.0                          | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-5.3.2                          | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |



| Amazon-EMR-Version<br>sbezeichnung | Tez-Version | Mit Tez installierte Komponent<br>en                                                                                                                                                                                                                                 |
|------------------------------------|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.3.1                          | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-5.3.0                          | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-5.2.3                          | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |

| Amazon-EMR-Version<br>sbezeichnung | Tez-Version | Mit Tez installierte Komponent<br>en                                                                                                                                                                                                                                 |
|------------------------------------|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.2.2                          | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-5.2.1                          | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-5.2.0                          | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |

| Amazon-EMR-Version<br>sbezeichnung | Tez-Version | Mit Tez installierte Komponent<br>en                                                                                                                                                                                                                                 |
|------------------------------------|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.1.1                          | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-5.1.0                          | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-5.0.3                          | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |

| Amazon-EMR-Version<br>sbezeichnung | Tez-Version | Mit Tez installierte Komponent<br>en                                                                                                                                                                                                                                 |
|------------------------------------|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.0.2                          | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-5.0.1                          | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-5.0.0                          | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |

| Amazon-EMR-Version<br>sbezeichnung | Tez-Version | Mit Tez installierte Komponent<br>en                                                                                                                                                                                                                                 |
|------------------------------------|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-4.9.6                          | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-4.9.5                          | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-4.9.4                          | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |

| Amazon-EMR-Version<br>sbezeichnung | Tez-Version | Mit Tez installierte Komponent<br>en                                                                                                                                                                                                                                 |
|------------------------------------|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-4.9.3                          | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-4.9.2                          | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-4.9.1                          | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |

| Amazon-EMR-Version<br>sbezeichnung | Tez-Version | Mit Tez installierte Komponent<br>en                                                                                                                                                                                                                                 |
|------------------------------------|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-4.8.5                          | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-4.8.4                          | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-4.8.3                          | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |

| Amazon-EMR-Version<br>sbezeichnung | Tez-Version | Mit Tez installierte Komponent<br>en                                                                                                                                                                                                                                 |
|------------------------------------|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-4.8.2                          | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-4.8.1                          | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-4.8.0                          | 0.8.4       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |



| Amazon-EMR-Version<br>sbezeichnung | Tez-Version | Mit Tez installierte Komponent<br>en                                                                                                                                                                                                                                 |
|------------------------------------|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-4.7.4                          | 0.8.3       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-4.7.3                          | 0.8.3       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-4.7.2                          | 0.8.3       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |

| Amazon-EMR-Version<br>sbezeichnung | Tez-Version | Mit Tez installierte Komponent<br>en                                                                                                                                                                                                                                 |
|------------------------------------|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-4.7.1                          | 0.8.3       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |
| emr-4.7.0                          | 0.8.3       | emrfs, emr-goodies, hadoop-cl<br>ient, hadoop-mapred, hadoop-<br>hdfs-datanode, hadoop-<br>hdfs-library, hadoop-hdfs-<br>namenode, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, tez-on-yarn |

## Tez-Versionshinweise nach Version

### Themen

- [Amazon EMR 7.1.0 — Tez-Versionshinweise](#)
- [Amazon EMR 6.15.0 – Tez-Versionshinweise](#)
- [Amazon EMR 6.14.0 – Tez-Versionshinweise](#)
- [Amazon EMR 6.13.0 – Tez-Versionshinweise](#)
- [Amazon EMR 6.12.0 – Tez-Versionshinweise](#)
- [Amazon EMR 6.11.0 – Tez-Versionshinweise](#)
- [Amazon EMR 6.10.0 – Tez-Versionshinweise](#)
- [Amazon EMR 6.9.0 – Tez-Versionshinweise](#)

- [Amazon EMR 6.8.0 – Tez-Versionshinweise](#)
- [Amazon EMR 6.7.0 – Tez-Versionshinweise](#)
- [Amazon EMR 6.6.0 – Tez-Versionshinweise](#)

## Amazon EMR 7.1.0 — Tez-Versionshinweise

### Amazon EMR 7.1.0 — Tez-Änderungen

| Typ            | Beschreibung                                                                            |
|----------------|-----------------------------------------------------------------------------------------|
| Fehlerbehebung | <a href="#">TEZ-4394</a> — Netty4 ShuffleHandler: sollte 1 Boss-Thread verwenden (#256) |
| Upgrade        | Führt ein Upgrade von Netty auf 4.1.100.Final durch.                                    |
| Upgrade        | Führt ein Upgrade von Jetty auf 9.4.53.v20231009 durch.                                 |

## Amazon EMR 6.15.0 – Tez-Versionshinweise

### Amazon EMR 6.15.0 – Tez-Änderungen

| Typ      | Beschreibung                                                     |
|----------|------------------------------------------------------------------|
| Funktion | <a href="#">TEZ-4397</a> : Tez-Eingabesplits asynchron öffnen    |
| Upgrade  | <a href="#">TEZ-4493</a> : Apache Hadoop auf 3.3.6 aktualisieren |

### Amazon EMR 6.15.0 – Tez-Features

- [Asynchrones Öffnen von Tez-Splits](#) – Tez in Amazon EMR 6.15.0 führt Konfigurationen ein, die Sie angeben können, um die Eingabe-Splits in einem gruppierten Tez-Split asynchron zu öffnen. Das Feature wurde von [TEZ-4397](#) initiiert, wies jedoch in OSS Hive Regressionen auf. Amazon EMR Hive hat die Regressionen und zusätzliche Fehler in der Hive-ACID-Tabelle behoben.

Diese Verbesserung führt zu einer schnelleren Leistung von Leseabfragen, wenn ein einzelner gruppierter Tez-Split eine große Anzahl von Eingabe-Splits enthält. Weitere Informationen finden Sie unter [Asynchrone Öffnung von Tez-Splits](#).

## Amazon EMR 6.14.0 – Tez-Versionshinweise

### Amazon EMR 6.14.0 – Änderungen bei Tez

| Typ          | Beschreibung                                     |
|--------------|--------------------------------------------------|
| Verbesserung | Aktualisieren Sie die TLS-Version in Tez auf 1.2 |

## Amazon EMR 6.13.0 – Tez-Versionshinweise

### Amazon EMR 6.13.0 – Änderungen bei Tez

| Typ            | Beschreibung                                                                                                                                                                          |
|----------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Fehlerbehebung | REVERT <a href="#">TEZ-4295</a> : Daten konnten nicht dekomprimiert werden. Die Pufferlänge ist zu klein.                                                                             |
| Fehlerbehebung | REVERT <a href="#">TEZ-4302</a> : <code>NullPointerException</code> in <code>CodecUtils</code> mit <code>GzipCodec</code> .                                                           |
| Fehlerbehebung | REVERT <a href="#">TEZ-4234</a> : Der Kompressor kann <code>IllegalArgumentException</code> in <code>Buffer.limit</code> , wo der Grenzwert die Kapazität überschreitet, verursachen. |
| Fehlerbehebung | REVERT <a href="#">TEZ-4135</a> : Verbessert die Speicherzuweisung bei der Ausführung von Lesevorgängen im Speicher.                                                                  |

## Amazon EMR 6.12.0 – Tez-Versionshinweise

### Amazon EMR 6.12.0 – Änderungen bei Tez

| Typ            | Beschreibung                                                                                                                                                                 |
|----------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Verbesserung   | Support für JDK-11- und JDK-17-Laufzeit hinzugefügt                                                                                                                          |
| Fehlerbehebung | <a href="#">TEZ-4492</a> : Aktualisieren Sie Bowerrc, um den bower.herokuapp mirror zu verwenden, um das CERT_EXPIRE-Problem in der Bower-Registry zu vermeiden (BOWER-2608) |
| Upgrade        | Surefire wurde auf 3.0.0-M7 aktualisiert                                                                                                                                     |

## Amazon EMR 6.11.0 – Tez-Versionshinweise

### Amazon EMR 6.11.0 – Änderungen bei Tez

| Typ          | Beschreibung                                                                                                                      |
|--------------|-----------------------------------------------------------------------------------------------------------------------------------|
| Fehler       | Ein ungültiger Übergang zum Scheitelpunktstatus bei der Bereinigung von Shuffle-Daten auf Scheitelpunktebene wurde behoben        |
| Fehler       | Es wurde behoben, dass die Bereinigung von Shuffle-Daten auf DAG- oder Vertexebe nicht funktionierte                              |
| Verbesserung | Aktivieren Sie standardmäßig <code>tez.am.dag.cleanup.on.completion</code> , um die Shuffle-Daten abgeschlossener DAGs zu löschen |

## Amazon EMR 6.10.0 – Tez-Versionshinweise

### Amazon EMR 6.10.0 – Tez-Änderungen

| Typ      | Beschreibung                                                                                                                                             |
|----------|----------------------------------------------------------------------------------------------------------------------------------------------------------|
| Funktion | <code>tez.runtime.transfer.data-via-events.enabled</code> standardmäßig aktivieren                                                                       |
| Backport | <a href="#">TEZ-4450</a> : Behebung eines Fehlers beim Abrufen von Shuffle-Daten, wenn Shuffle-Daten über Datenverschiebungsereignisse übertragen wurden |
| Backport | <a href="#">TEZ-4460</a> : Der Lese-Timeout-Fehler beim Abrufen von Shuffle-Daten aus dem Tez Shuffle Handler wurde behoben                              |
| Backport | <a href="#">TEZ-4455: Pipeline für bessere Debuggbarkeit hinzugefügt</a> LoggingHandler ShuffleHandler                                                   |
| Fehler   | Behebt, dass die Tez-Aufgabe zeitweise hängen bleibt, wenn das Preemptive der Aufgabe aktiviert ist                                                      |

## Amazon EMR 6.9.0 – Tez-Versionshinweise

### Amazon EMR 6.9.0 – Tez-Änderungen

| Typ     | Beschreibung                                                                                                                      |
|---------|-----------------------------------------------------------------------------------------------------------------------------------|
| Upgrade | Tez wurde auf 0.10.2 aktualisiert. Weitere Informationen finden Sie im <a href="#">Änderungsprotokoll für Apache Tez 0.10.2</a> . |
| Upgrade | Upgrade auf Hadoop 3.3.3.                                                                                                         |

| Typ    | Beschreibung                                                                                                               |
|--------|----------------------------------------------------------------------------------------------------------------------------|
| Fehler | <code>tez.runtime.transfer.data-via-events.enabled</code> aufgrund von <a href="#">TEZ-4450</a> standardmäßig deaktiviert. |

## Amazon EMR 6.8.0 – Tez-Versionshinweise

### Amazon EMR 6.8.0 – Änderungen bei Tez

| Typ      | Beschreibung                                                                                                                |
|----------|-----------------------------------------------------------------------------------------------------------------------------|
| Backport | <a href="#">TEZ-3363</a> : Löscht Zwischendaten auf Scheitelpunkteebene für Shuffle Handler                                 |
| Backport | <a href="#">TEZ-4129</a> : Daten für Zwischenversuche bei fehlgeschlagenen Versuchen für Shuffle Handler löschen            |
| Backport | <a href="#">TEZ-4430</a> : Es wurde behoben, dass die Eigenschaft <code>tez.task.launch.cmd-opts</code> nicht funktionierte |

## Amazon EMR 6.7.0 – Tez-Versionshinweise

### Amazon EMR 6.7.0 – Tez-Änderungen

| Typ      | Beschreibung                                                                               |
|----------|--------------------------------------------------------------------------------------------|
| Backport | <a href="#">TEZ-4403</a> : Aktualisieren Sie die SLF4J-Version auf 1.7.36                  |
| Backport | <a href="#">TEZ-4405</a> : Ersetzen Sie <code>log4j 1.x</code> durch <code>reload4j</code> |
| Backport | <a href="#">TEZ-4411: Tez-Build-Fehler: .js nicht gefunden</a><br>FileSaver                |

## Amazon EMR 6.6.0 – Tez-Versionshinweise

### Amazon EMR 6.6.0 – Tez-Änderungen

| Typ      | Beschreibung                                                                                                          |
|----------|-----------------------------------------------------------------------------------------------------------------------|
| Backport | <a href="#">TEZ-3918</a> : Es wurde behoben, dass die Eigenschaft <code>tez.task.log.level</code> nicht funktioniert. |
| Backport | <a href="#">TEZ-4353</a> : Aktualisieren Sie <code>commons-io</code> auf 2.8.0.                                       |
| Backport | <a href="#">TEZ-4114</a> : Entfernt die direkte Jetty-Abhängigkeit von TEZ.                                           |
| Backport | <a href="#">TEZ-4323</a> : Jetty Jars wurden zusammen mit TEZ-4114 aus dem Paket <code>dist</code> entfernt.          |



# Apache Zeppelin

Verwenden Sie Apache Zeppelin als Notizbuch für eine interaktive Datenauswertung. Weitere Informationen zu Zeppelin finden Sie unter <https://zeppelin.apache.org/>. Zeppelin ist in Amazon EMR Versionen 5.0.0 und höher enthalten. Frühere Versionen enthalten Zeppelin als Sandbox-Anwendung. Weitere Informationen finden Sie unter [Amazon-EMR-4.x-Versionen](#).

Richten Sie für den Zugriff auf die Zeppelin-Webschnittstelle einen SSH-Tunnel zum Master-Knoten und eine Proxy-Verbindung ein. Weitere Informationen finden Sie unter [Auf Amazon-EMR-Clustern gehostete Webschnittstellen anzeigen](#).

Die folgende Tabelle listet die Version von Zeppelin auf, die in der neuesten Version der Amazon-EMR-7.x-Serie enthalten ist, zusammen mit den Komponenten, die Amazon EMR mit Zeppelin installiert.

Die Version der Komponenten, die in dieser Version mit Zeppelin installiert wurden, finden Sie unter [Komponentenversionen von Version 7.1.0](#).

Zeppelin-Versionsinformationen für emr-7.1.0

| Amazon-EMR-Version<br>sbezeichnung | Zeppelin-Version | Mit Zeppelin installierte<br>Komponenten                                                                                                                                                                                                                                                                                                              |
|------------------------------------|------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-7.1.0                          | Zeppelin 0.10.1  | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |

Die folgende Tabelle listet die Version von Zeppelin auf, die in der neuesten Version der Amazon-EMR-6.x-Serie enthalten ist, zusammen mit den Komponenten, die Amazon EMR mit Zeppelin installiert.

Die Version der Komponenten, die mit Zeppelin in dieser Version installiert wurden, finden Sie unter [Komponentenversionen der Version 6.15.0](#).

#### Zeppelin-Versionsinformationen für emr-6.15.0

| Amazon-EMR-Version<br>sbezeichnung | Zeppelin-Version | Mit Zeppelin installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                        |
|------------------------------------|------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.15.0                         | Zeppelin 0.10.1  | aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |

Die folgende Tabelle listet die Version von Zeppelin auf, die in der neuesten Version der Amazon-EMR-5.x-Serie enthalten ist, zusammen mit den Komponenten, die Amazon EMR mit Zeppelin installiert.

[Informationen zur Version der Komponenten, die in dieser Version mit Zeppelin installiert wurden, finden Sie unter Komponentenversionen von Version 5.36.2.](#)

## Zeppelin-Versionsinformationen für emr-5.36.2

| Amazon-EMR-Version<br>sbezeichnung | Zeppelin-Version | Mit Zeppelin installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                                                                           |
|------------------------------------|------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.36.2                         | Zeppelin 0.10.0  | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema-<br>nager, hadoop-yarn-resour-<br>cemanager, hadoop-yarn-<br>timeline-server, hudi, hudi-<br>spark, livy-server, r, spark-<br>client, spark-history-server,<br>spark-on-yarn, spark-yarn-<br>slave, zeppelin-server |

### Themen

- [Überlegungen zur Verwendung von Zeppelin in Amazon EMR](#)
- [Zeppelin-Versionshistorie](#)

## Überlegungen zur Verwendung von Zeppelin in Amazon EMR

- Stellen Sie eine Verbindung mit Zeppelin über dieselbe [SSH-Tunnel-Methode](#) her, die Sie auch für andere Webserver auf dem Master-Knoten verwenden. Der Zeppelin-Server befindet sich an Port 8890.
- Zeppelin für Amazon EMR ab Version 5.0.0 unterstützt die [Shiro-Authentifizierung](#).
- Zeppelin auf Amazon EMR, Version 5.8.0 und höher, unterstützt die Verwendung von AWS Glue Data Catalog als Metastore für Spark SQL. Weitere Informationen finden Sie unter [Verwenden des AWS Glue-Datenkatalogs als Metastore für Spark SQL](#).
- Zeppelin verwendet einige der in der Konfigurationsdatei `spark-defaults.conf` Ihres Clusters definierten Einstellungen nicht, auch wenn YARN angewiesen wird, Executor dynamisch

zuzuweisen, wenn Sie `spark.dynamicAllocation.enabled` auf `true` gesetzt haben. Sie müssen Executor-Einstellungen, z. B. Arbeitsspeicher und Prozessorkerne, über die Zeppelin-Registerkarte Interpreter festlegen. Starten Sie dann den Interpreter neu, um die Änderungen zu übernehmen.

- Amazon-EMR-Versionen 6.10.0 und höher unterstützen die Apache-Zeppelin-Integration mit Apache Flink. Weitere Informationen finden Sie unter [Arbeiten mit Flink-Aufträgen von Zeppelin in Amazon EMR](#).
- Zeppelin in Amazon EMR unterstützt den SparkR-Interpreter nicht.

## Zeppelin-Versionshistorie

In der folgenden Tabelle sind die Version von Zeppelin aufgeführt, die in jeder Release-Version von Amazon EMR enthalten ist, zusammen mit den Komponenten, die mit der Anwendung installiert wurden. Informationen zu den Komponentenversionen in den einzelnen Versionen finden Sie im Abschnitt [Komponentenversion für Ihre Version in Amazon-EMR-7.x-Versionen](#), [Amazon-EMR-6.x-Versionen](#) oder [Amazon-EMR-5.x-Versionen](#).

### Zeppelin-Versionsinformationen

| Amazon-EMR-Version<br>sbezeichnung | Zeppelin-Version | Mit Zeppelin installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                       |
|------------------------------------|------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.36.2                         | 0.10.0           | aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |

| Amazon-EMR-Version<br>sbezeichnung | Zeppelin-Version | Mit Zeppelin installierte<br>Komponenten                                                                                                                                                                                                                                                                                                               |
|------------------------------------|------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-7.1.0                          | 0.10.1           | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |
| emr-7.0.0                          | 0.10.1           | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |

| Amazon-EMR-Version<br>sbezeichnung | Zeppelin-Version | Mit Zeppelin installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                                                                         |
|------------------------------------|------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.15.0                         | 0.10.1           | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, hudi, hudi-<br>spark, livy-server, r, spark-<br>client, spark-history-server,<br>spark-on-yarn, spark-yarn-<br>slave, zeppelin-server |
| emr-6.14.0                         | 0.10.1           | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, hudi, hudi-<br>spark, livy-server, r, spark-<br>client, spark-history-server,<br>spark-on-yarn, spark-yarn-<br>slave, zeppelin-server |

| Amazon-EMR-Version<br>sbezeichnung | Zeppelin-Version | Mit Zeppelin installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                                                                         |
|------------------------------------|------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.13.0                         | 0.10.1           | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, hudi, hudi-<br>spark, livy-server, r, spark-<br>client, spark-history-server,<br>spark-on-yarn, spark-yarn-<br>slave, zeppelin-server |
| emr-6.12.0                         | 0.10.1           | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, hudi, hudi-<br>spark, livy-server, r, spark-<br>client, spark-history-server,<br>spark-on-yarn, spark-yarn-<br>slave, zeppelin-server |

| Amazon-EMR-Version<br>sbezeichnung | Zeppelin-Version | Mit Zeppelin installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                        |
|------------------------------------|------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.11.1                         | 0.10.1           | aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |
| emr-6.11.0                         | 0.10.1           | aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |



| Amazon-EMR-Version<br>sbezeichnung | Zeppelin-Version | Mit Zeppelin installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                                                                           |
|------------------------------------|------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.10.1                         | 0.10.1           | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema-<br>nager, hadoop-yarn-resour-<br>cemanager, hadoop-yarn-<br>timeline-server, hudi, hudi-<br>spark, livy-server, r, spark-<br>client, spark-history-server,<br>spark-on-yarn, spark-yarn-<br>slave, zeppelin-server |
| emr-6.10.0                         | 0.10.1           | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema-<br>nager, hadoop-yarn-resour-<br>cemanager, hadoop-yarn-<br>timeline-server, hudi, hudi-<br>spark, livy-server, r, spark-<br>client, spark-history-server,<br>spark-on-yarn, spark-yarn-<br>slave, zeppelin-server |

| Amazon-EMR-Version<br>sbezeichnung | Zeppelin-Version | Mit Zeppelin installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                                                                         |
|------------------------------------|------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.9.1                          | 0.10.1           | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, hudi, hudi-<br>spark, livy-server, r, spark-<br>client, spark-history-server,<br>spark-on-yarn, spark-yarn-<br>slave, zeppelin-server |
| emr-6.9.0                          | 0.10.1           | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, hudi, hudi-<br>spark, livy-server, r, spark-<br>client, spark-history-server,<br>spark-on-yarn, spark-yarn-<br>slave, zeppelin-server |

| Amazon-EMR-Version<br>sbezeichnung | Zeppelin-Version | Mit Zeppelin installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                                                                         |
|------------------------------------|------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.8.1                          | 0.10.1           | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, hudi, hudi-<br>spark, livy-server, r, spark-<br>client, spark-history-server,<br>spark-on-yarn, spark-yarn-<br>slave, zeppelin-server |
| emr-6.8.0                          | 0.10.1           | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, hudi, hudi-<br>spark, livy-server, r, spark-<br>client, spark-history-server,<br>spark-on-yarn, spark-yarn-<br>slave, zeppelin-server |

| Amazon-EMR-Version<br>sbezeichnung | Zeppelin-Version | Mit Zeppelin installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                                                                         |
|------------------------------------|------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.7.0                          | 0.10.0           | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, hudi, hudi-<br>spark, livy-server, r, spark-<br>client, spark-history-server,<br>spark-on-yarn, spark-yarn-<br>slave, zeppelin-server |
| emr-5.36.1                         | 0.10.0           | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, hudi, hudi-<br>spark, livy-server, r, spark-<br>client, spark-history-server,<br>spark-on-yarn, spark-yarn-<br>slave, zeppelin-server |

| Amazon-EMR-Version<br>sbezeichnung | Zeppelin-Version | Mit Zeppelin installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                                                                         |
|------------------------------------|------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.36.0                         | 0.10.0           | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, hudi, hudi-<br>spark, livy-server, r, spark-<br>client, spark-history-server,<br>spark-on-yarn, spark-yarn-<br>slave, zeppelin-server |
| emr-6.6.0                          | 0.10.0           | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, hudi, hudi-<br>spark, livy-server, r, spark-<br>client, spark-history-server,<br>spark-on-yarn, spark-yarn-<br>slave, zeppelin-server |

| Amazon-EMR-Version<br>sbezeichnung | Zeppelin-Version | Mit Zeppelin installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                                                      |
|------------------------------------|------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.35.0                         | 0.10.0           | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema-<br>nager, hadoop-yarn-resour-<br>cemanager, hadoop-yarn-<br>timeline-server, livy-server, r,<br>spark-client, spark-history-serv-<br>er, spark-on-yarn, spark-yarn-<br>slave, zeppelin-server |
| emr-6.5.0                          | 0.10.0           | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema-<br>nager, hadoop-yarn-resour-<br>cemanager, hadoop-yarn-<br>timeline-server, livy-server, r,<br>spark-client, spark-history-serv-<br>er, spark-on-yarn, spark-yarn-<br>slave, zeppelin-server |

| Amazon-EMR-Version<br>sbezeichnung | Zeppelin-Version | Mit Zeppelin installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                      |
|------------------------------------|------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.4.0                          | 0.9.0            | aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |
| emr-6.3.1                          | 0.9.0            | aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |

| Amazon-EMR-Version<br>sbezeichnung | Zeppelin-Version | Mit Zeppelin installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                      |
|------------------------------------|------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.3.0                          | 0.9.0            | aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |
| emr-6.2.1                          | 0.9.0            | aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |



| Amazon-EMR-Version<br>sbezeichnung | Zeppelin-Version | Mit Zeppelin installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                                                      |
|------------------------------------|------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.2.0                          | 0.9.0            | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema-<br>nager, hadoop-yarn-resour-<br>cemanager, hadoop-yarn-<br>timeline-server, livy-server, r,<br>spark-client, spark-history-serv-<br>er, spark-on-yarn, spark-yarn-<br>slave, zeppelin-server |
| emr-6.1.1                          | 0.9.0            | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema-<br>nager, hadoop-yarn-resour-<br>cemanager, hadoop-yarn-<br>timeline-server, livy-server, r,<br>spark-client, spark-history-serv-<br>er, spark-on-yarn, spark-yarn-<br>slave, zeppelin-server |

| Amazon-EMR-Version<br>sbezeichnung | Zeppelin-Version | Mit Zeppelin installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                                                      |
|------------------------------------|------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.1.0                          | 0.9.0            | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema-<br>nager, hadoop-yarn-resour-<br>cemanager, hadoop-yarn-<br>timeline-server, livy-server, r,<br>spark-client, spark-history-serv-<br>er, spark-on-yarn, spark-yarn-<br>slave, zeppelin-server |
| emr-6.0.1                          | 0.9.0            | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema-<br>nager, hadoop-yarn-resour-<br>cemanager, hadoop-yarn-<br>timeline-server, livy-server, r,<br>spark-client, spark-history-serv-<br>er, spark-on-yarn, spark-yarn-<br>slave, zeppelin-server |

| Amazon-EMR-Version<br>sbezeichnung | Zeppelin-Version | Mit Zeppelin installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                                                      |
|------------------------------------|------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.0.0                          | 0.9.0            | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema-<br>nager, hadoop-yarn-resour-<br>cemanager, hadoop-yarn-<br>timeline-server, livy-server, r,<br>spark-client, spark-history-serv-<br>er, spark-on-yarn, spark-yarn-<br>slave, zeppelin-server |
| emr-5.34.0                         | 0.10.0           | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema-<br>nager, hadoop-yarn-resour-<br>cemanager, hadoop-yarn-<br>timeline-server, livy-server, r,<br>spark-client, spark-history-serv-<br>er, spark-on-yarn, spark-yarn-<br>slave, zeppelin-server |

| Amazon-EMR-Version<br>sbezeichnung | Zeppelin-Version | Mit Zeppelin installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                                                      |
|------------------------------------|------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.33.1                         | 0.9.0            | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema-<br>nager, hadoop-yarn-resour-<br>cemanager, hadoop-yarn-<br>timeline-server, livy-server, r,<br>spark-client, spark-history-serv-<br>er, spark-on-yarn, spark-yarn-<br>slave, zeppelin-server |
| emr-5.33.0                         | 0.9.0            | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema-<br>nager, hadoop-yarn-resour-<br>cemanager, hadoop-yarn-<br>timeline-server, livy-server, r,<br>spark-client, spark-history-serv-<br>er, spark-on-yarn, spark-yarn-<br>slave, zeppelin-server |

| Amazon-EMR-Version<br>sbezeichnung | Zeppelin-Version | Mit Zeppelin installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                     |
|------------------------------------|------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.32.1                         | 0.8.2            | aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |
| emr-5.32.0                         | 0.8.2            | aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |

| Amazon-EMR-Version<br>sbezeichnung | Zeppelin-Version | Mit Zeppelin installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                                                      |
|------------------------------------|------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.31.1                         | 0.8.2            | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema-<br>nager, hadoop-yarn-resour-<br>cemanager, hadoop-yarn-<br>timeline-server, livy-server, r,<br>spark-client, spark-history-serv-<br>er, spark-on-yarn, spark-yarn-<br>slave, zeppelin-server |
| emr-5.31.0                         | 0.8.2            | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema-<br>nager, hadoop-yarn-resour-<br>cemanager, hadoop-yarn-<br>timeline-server, livy-server, r,<br>spark-client, spark-history-serv-<br>er, spark-on-yarn, spark-yarn-<br>slave, zeppelin-server |

| Amazon-EMR-Version<br>sbezeichnung | Zeppelin-Version | Mit Zeppelin installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                      |
|------------------------------------|------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.30.2                         | 0.8.2            | aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |
| emr-5.30.1                         | 0.8.2            | aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |

| Amazon-EMR-Version<br>sbezeichnung | Zeppelin-Version | Mit Zeppelin installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                                                   |
|------------------------------------|------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.30.0                         | 0.8.2            | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, livy-server, r,<br>spark-client, spark-history-serv<br>er, spark-on-yarn, spark-yarn-<br>slave, zeppelin-server |
| emr-5.29.0                         | 0.8.2            | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, livy-server, r,<br>spark-client, spark-history-serv<br>er, spark-on-yarn, spark-yarn-<br>slave, zeppelin-server |



| Amazon-EMR-Version<br>sbezeichnung | Zeppelin-Version | Mit Zeppelin installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                      |
|------------------------------------|------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.28.1                         | 0.8.2            | aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |
| emr-5.28.0                         | 0.8.2            | aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |

| Amazon-EMR-Version<br>sbezeichnung | Zeppelin-Version | Mit Zeppelin installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                      |
|------------------------------------|------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.27.1                         | 0.8.1            | aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |
| emr-5.27.0                         | 0.8.1            | aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |

| Amazon-EMR-Version<br>sbezeichnung | Zeppelin-Version | Mit Zeppelin installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                                                      |
|------------------------------------|------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.26.0                         | 0.8.1            | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema-<br>nager, hadoop-yarn-resour-<br>cemanager, hadoop-yarn-<br>timeline-server, livy-server, r,<br>spark-client, spark-history-serv-<br>er, spark-on-yarn, spark-yarn-<br>slave, zeppelin-server |
| emr-5.25.0                         | 0.8.1            | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema-<br>nager, hadoop-yarn-resour-<br>cemanager, hadoop-yarn-<br>timeline-server, livy-server, r,<br>spark-client, spark-history-serv-<br>er, spark-on-yarn, spark-yarn-<br>slave, zeppelin-server |

| Amazon-EMR-Version<br>sbezeichnung | Zeppelin-Version | Mit Zeppelin installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                                                      |
|------------------------------------|------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.24.1                         | 0.8.1            | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema-<br>nager, hadoop-yarn-resour-<br>cemanager, hadoop-yarn-<br>timeline-server, livy-server, r,<br>spark-client, spark-history-serv-<br>er, spark-on-yarn, spark-yarn-<br>slave, zeppelin-server |
| emr-5.24.0                         | 0.8.1            | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema-<br>nager, hadoop-yarn-resour-<br>cemanager, hadoop-yarn-<br>timeline-server, livy-server, r,<br>spark-client, spark-history-serv-<br>er, spark-on-yarn, spark-yarn-<br>slave, zeppelin-server |

| Amazon-EMR-Version<br>sbezeichnung | Zeppelin-Version | Mit Zeppelin installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                      |
|------------------------------------|------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.23.1                         | 0.8.1            | aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |
| emr-5.23.0                         | 0.8.1            | aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |

| Amazon-EMR-Version<br>sbezeichnung | Zeppelin-Version | Mit Zeppelin installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                                                      |
|------------------------------------|------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.22.0                         | 0.8.1            | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema-<br>nager, hadoop-yarn-resour-<br>cemanager, hadoop-yarn-<br>timeline-server, livy-server, r,<br>spark-client, spark-history-serv-<br>er, spark-on-yarn, spark-yarn-<br>slave, zeppelin-server |
| emr-5.21.2                         | 0.8.0            | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema-<br>nager, hadoop-yarn-resour-<br>cemanager, hadoop-yarn-<br>timeline-server, r, spark-client,<br>spark-history-server, spark-<br>on-yarn, spark-yarn-slave,<br>zeppelin-server                |

| Amazon-EMR-Version<br>sbezeichnung | Zeppelin-Version | Mit Zeppelin installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                                       |
|------------------------------------|------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.21.1                         | 0.8.0            | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema-<br>nager, hadoop-yarn-resour-<br>cemanager, hadoop-yarn-<br>timeline-server, r, spark-client,<br>spark-history-server, spark-<br>on-yarn, spark-yarn-slave,<br>zeppelin-server |
| emr-5.21.0                         | 0.8.0            | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema-<br>nager, hadoop-yarn-resour-<br>cemanager, hadoop-yarn-<br>timeline-server, r, spark-client,<br>spark-history-server, spark-<br>on-yarn, spark-yarn-slave,<br>zeppelin-server |

| Amazon-EMR-Version<br>sbezeichnung | Zeppelin-Version | Mit Zeppelin installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                                       |
|------------------------------------|------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.20.1                         | 0.8.0            | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema-<br>nager, hadoop-yarn-resour-<br>cemanager, hadoop-yarn-<br>timeline-server, r, spark-client,<br>spark-history-server, spark-<br>on-yarn, spark-yarn-slave,<br>zeppelin-server |
| emr-5.20.0                         | 0.8.0            | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema-<br>nager, hadoop-yarn-resour-<br>cemanager, hadoop-yarn-<br>timeline-server, r, spark-client,<br>spark-history-server, spark-<br>on-yarn, spark-yarn-slave,<br>zeppelin-server |



| Amazon-EMR-Version<br>sbezeichnung | Zeppelin-Version | Mit Zeppelin installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                                       |
|------------------------------------|------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.19.1                         | 0.8.0            | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema-<br>nager, hadoop-yarn-resour-<br>cemanager, hadoop-yarn-<br>timeline-server, r, spark-client,<br>spark-history-server, spark-<br>on-yarn, spark-yarn-slave,<br>zeppelin-server |
| emr-5.19.0                         | 0.8.0            | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema-<br>nager, hadoop-yarn-resour-<br>cemanager, hadoop-yarn-<br>timeline-server, r, spark-client,<br>spark-history-server, spark-<br>on-yarn, spark-yarn-slave,<br>zeppelin-server |

| Amazon-EMR-Version<br>sbezeichnung | Zeppelin-Version | Mit Zeppelin installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                                       |
|------------------------------------|------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.18.1                         | 0.8.0            | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema-<br>nager, hadoop-yarn-resour-<br>cemanager, hadoop-yarn-<br>timeline-server, r, spark-client,<br>spark-history-server, spark-<br>on-yarn, spark-yarn-slave,<br>zeppelin-server |
| emr-5.18.0                         | 0.8.0            | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema-<br>nager, hadoop-yarn-resour-<br>cemanager, hadoop-yarn-<br>timeline-server, r, spark-client,<br>spark-history-server, spark-<br>on-yarn, spark-yarn-slave,<br>zeppelin-server |

| Amazon-EMR-Version<br>sbezeichnung | Zeppelin-Version | Mit Zeppelin installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                                       |
|------------------------------------|------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.17.2                         | 0.7.3            | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema-<br>nager, hadoop-yarn-resour-<br>cemanager, hadoop-yarn-<br>timeline-server, r, spark-client,<br>spark-history-server, spark-<br>on-yarn, spark-yarn-slave,<br>zeppelin-server |
| emr-5.17.1                         | 0.7.3            | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema-<br>nager, hadoop-yarn-resour-<br>cemanager, hadoop-yarn-<br>timeline-server, r, spark-client,<br>spark-history-server, spark-<br>on-yarn, spark-yarn-slave,<br>zeppelin-server |

| Amazon-EMR-Version<br>sbezeichnung | Zeppelin-Version | Mit Zeppelin installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                                       |
|------------------------------------|------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.17.0                         | 0.7.3            | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema-<br>nager, hadoop-yarn-resour-<br>cemanager, hadoop-yarn-<br>timeline-server, r, spark-client,<br>spark-history-server, spark-<br>on-yarn, spark-yarn-slave,<br>zeppelin-server |
| emr-5.16.1                         | 0.7.3            | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema-<br>nager, hadoop-yarn-resour-<br>cemanager, hadoop-yarn-<br>timeline-server, r, spark-client,<br>spark-history-server, spark-<br>on-yarn, spark-yarn-slave,<br>zeppelin-server |

| Amazon-EMR-Version<br>sbezeichnung | Zeppelin-Version | Mit Zeppelin installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                                       |
|------------------------------------|------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.16.0                         | 0.7.3            | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema-<br>nager, hadoop-yarn-resour-<br>cemanager, hadoop-yarn-<br>timeline-server, r, spark-client,<br>spark-history-server, spark-<br>on-yarn, spark-yarn-slave,<br>zeppelin-server |
| emr-5.15.1                         | 0.7.3            | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema-<br>nager, hadoop-yarn-resour-<br>cemanager, hadoop-yarn-<br>timeline-server, r, spark-client,<br>spark-history-server, spark-<br>on-yarn, spark-yarn-slave,<br>zeppelin-server |

| Amazon-EMR-Version<br>sbezeichnung | Zeppelin-Version | Mit Zeppelin installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                                       |
|------------------------------------|------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.15.0                         | 0.7.3            | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema-<br>nager, hadoop-yarn-resour-<br>cemanager, hadoop-yarn-<br>timeline-server, r, spark-client,<br>spark-history-server, spark-<br>on-yarn, spark-yarn-slave,<br>zeppelin-server |
| emr-5.14.2                         | 0.7.3            | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema-<br>nager, hadoop-yarn-resour-<br>cemanager, hadoop-yarn-<br>timeline-server, r, spark-client,<br>spark-history-server, spark-<br>on-yarn, spark-yarn-slave,<br>zeppelin-server |

| Amazon-EMR-Version<br>sbezeichnung | Zeppelin-Version | Mit Zeppelin installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                                       |
|------------------------------------|------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.14.1                         | 0.7.3            | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema-<br>nager, hadoop-yarn-resour-<br>cemanager, hadoop-yarn-<br>timeline-server, r, spark-client,<br>spark-history-server, spark-<br>on-yarn, spark-yarn-slave,<br>zeppelin-server |
| emr-5.14.0                         | 0.7.3            | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema-<br>nager, hadoop-yarn-resour-<br>cemanager, hadoop-yarn-<br>timeline-server, r, spark-client,<br>spark-history-server, spark-<br>on-yarn, spark-yarn-slave,<br>zeppelin-server |

| Amazon-EMR-Version<br>sbezeichnung | Zeppelin-Version | Mit Zeppelin installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                                       |
|------------------------------------|------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.13.1                         | 0.7.3            | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema-<br>nager, hadoop-yarn-resour-<br>cemanager, hadoop-yarn-<br>timeline-server, r, spark-client,<br>spark-history-server, spark-<br>on-yarn, spark-yarn-slave,<br>zeppelin-server |
| emr-5.13.0                         | 0.7.3            | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema-<br>nager, hadoop-yarn-resour-<br>cemanager, hadoop-yarn-<br>timeline-server, r, spark-client,<br>spark-history-server, spark-<br>on-yarn, spark-yarn-slave,<br>zeppelin-server |



| Amazon-EMR-Version<br>sbezeichnung | Zeppelin-Version | Mit Zeppelin installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                                  |
|------------------------------------|------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.12.3                         | 0.7.3            | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, spark-client,<br>spark-history-server, spark-<br>on-yarn, spark-yarn-slave,<br>zeppelin-server |
| emr-5.12.2                         | 0.7.3            | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, spark-client,<br>spark-history-server, spark-<br>on-yarn, spark-yarn-slave,<br>zeppelin-server |

| Amazon-EMR-Version<br>sbezeichnung | Zeppelin-Version | Mit Zeppelin installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                                  |
|------------------------------------|------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.12.1                         | 0.7.3            | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, spark-client,<br>spark-history-server, spark-<br>on-yarn, spark-yarn-slave,<br>zeppelin-server |
| emr-5.12.0                         | 0.7.3            | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, spark-client,<br>spark-history-server, spark-<br>on-yarn, spark-yarn-slave,<br>zeppelin-server |

| Amazon-EMR-Version<br>sbezeichnung | Zeppelin-Version | Mit Zeppelin installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                                    |
|------------------------------------|------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.11.4                         | 0.7.3            | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema-<br>nager, hadoop-yarn-resour-<br>cemanager, hadoop-yarn-<br>timeline-server, spark-client,<br>spark-history-server, spark-<br>on-yarn, spark-yarn-slave,<br>zeppelin-server |
| emr-5.11.3                         | 0.7.3            | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema-<br>nager, hadoop-yarn-resour-<br>cemanager, hadoop-yarn-<br>timeline-server, spark-client,<br>spark-history-server, spark-<br>on-yarn, spark-yarn-slave,<br>zeppelin-server |

| Amazon-EMR-Version<br>sbezeichnung | Zeppelin-Version | Mit Zeppelin installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                                  |
|------------------------------------|------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.11.2                         | 0.7.3            | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, spark-client,<br>spark-history-server, spark-<br>on-yarn, spark-yarn-slave,<br>zeppelin-server |
| emr-5.11.1                         | 0.7.3            | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, spark-client,<br>spark-history-server, spark-<br>on-yarn, spark-yarn-slave,<br>zeppelin-server |

| Amazon-EMR-Version<br>sbezeichnung | Zeppelin-Version | Mit Zeppelin installierte<br>Komponenten                                                                                                                                                                                                                                                                                                                                                  |
|------------------------------------|------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.11.0                         | 0.7.3            | aws-sagemaker-spark-sdk,<br>emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, spark-client,<br>spark-history-server, spark-<br>on-yarn, spark-yarn-slave,<br>zeppelin-server |
| emr-5.10.1                         | 0.7.3            | emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema<br>nager, hadoop-yarn-resour<br>cemanager, hadoop-yarn-<br>timeline-server, spark-client,<br>spark-history-server, spark-<br>on-yarn, spark-yarn-slave,<br>zeppelin-server                             |

| Amazon-EMR-Version<br>sbezeichnung | Zeppelin-Version | Mit Zeppelin installierte<br>Komponenten                                                                                                                                                                                                                                                                            |
|------------------------------------|------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.10.0                         | 0.7.3            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |
| emr-5.9.1                          | 0.7.2            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |

| Amazon-EMR-Version<br>sbezeichnung | Zeppelin-Version | Mit Zeppelin installierte<br>Komponenten                                                                                                                                                                                                                                                                            |
|------------------------------------|------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.9.0                          | 0.7.2            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |
| emr-5.8.3                          | 0.7.2            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |

| Amazon-EMR-Version<br>sbezeichnung | Zeppelin-Version | Mit Zeppelin installierte<br>Komponenten                                                                                                                                                                                                                                                                            |
|------------------------------------|------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.8.2                          | 0.7.2            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |
| emr-5.8.1                          | 0.7.2            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |



| Amazon-EMR-Version<br>sbezeichnung | Zeppelin-Version | Mit Zeppelin installierte<br>Komponenten                                                                                                                                                                                                                                                                            |
|------------------------------------|------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.8.0                          | 0.7.2            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |
| emr-5.7.1                          | 0.7.2            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |

| Amazon-EMR-Version<br>sbezeichnung | Zeppelin-Version | Mit Zeppelin installierte<br>Komponenten                                                                                                                                                                                                                                                                            |
|------------------------------------|------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.7.0                          | 0.7.2            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |
| emr-5.6.1                          | 0.7.1            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |

| Amazon-EMR-Version<br>sbezeichnung | Zeppelin-Version | Mit Zeppelin installierte<br>Komponenten                                                                                                                                                                                                                                                                            |
|------------------------------------|------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.6.0                          | 0.7.1            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |
| emr-5.5.4                          | 0.7.1            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server                              |

| Amazon-EMR-Version<br>sbezeichnung | Zeppelin-Version | Mit Zeppelin installierte<br>Komponenten                                                                                                                                                                                                                                                |
|------------------------------------|------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.5.3                          | 0.7.1            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |
| emr-5.5.2                          | 0.7.1            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |

| Amazon-EMR-Version<br>sbezeichnung | Zeppelin-Version | Mit Zeppelin installierte<br>Komponenten                                                                                                                                                                                                                                                |
|------------------------------------|------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.5.1                          | 0.7.1            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |
| emr-5.5.0                          | 0.7.1            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |

| Amazon-EMR-Version<br>sbezeichnung | Zeppelin-Version | Mit Zeppelin installierte<br>Komponenten                                                                                                                                                                                                                                                |
|------------------------------------|------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.4.1                          | 0.7.0            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |
| emr-5.4.0                          | 0.7.0            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |

| Amazon-EMR-Version<br>sbezeichnung | Zeppelin-Version | Mit Zeppelin installierte<br>Komponenten                                                                                                                                                                                                                                                |
|------------------------------------|------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.3.2                          | 0.6.2            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |
| emr-5.3.1                          | 0.6.2            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |

| Amazon-EMR-Version<br>sbezeichnung | Zeppelin-Version | Mit Zeppelin installierte<br>Komponenten                                                                                                                                                                                                                                                |
|------------------------------------|------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.3.0                          | 0.6.2            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |
| emr-5.2.3                          | 0.6.2            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |



| Amazon-EMR-Version<br>sbezeichnung | Zeppelin-Version | Mit Zeppelin installierte<br>Komponenten                                                                                                                                                                                                                                                |
|------------------------------------|------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.2.2                          | 0.6.2            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |
| emr-5.2.1                          | 0.6.2            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |

| Amazon-EMR-Version<br>sbezeichnung | Zeppelin-Version | Mit Zeppelin installierte<br>Komponenten                                                                                                                                                                                                                                                |
|------------------------------------|------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.2.0                          | 0.6.2            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |
| emr-5.1.1                          | 0.6.2            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |

| Amazon-EMR-Version<br>sbezeichnung | Zeppelin-Version | Mit Zeppelin installierte<br>Komponenten                                                                                                                                                                                                                                                |
|------------------------------------|------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.1.0                          | 0.6.2            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |
| emr-5.0.3                          | 0.6.1            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |

| Amazon-EMR-Version<br>sbezeichnung | Zeppelin-Version | Mit Zeppelin installierte<br>Komponenten                                                                                                                                                                                                                                                |
|------------------------------------|------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.0.2                          | 0.6.1            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |
| emr-5.0.1                          | 0.6.1            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |

| Amazon-EMR-Version<br>sbezeichnung | Zeppelin-Version | Mit Zeppelin installierte<br>Komponenten                                                                                                                                                                                                                                                |
|------------------------------------|------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.0.0                          | 0.6.1            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |

# Apache ZooKeeper

Apache ZooKeeper ist ein zentraler Dienst zur Verwaltung von Konfigurationsinformationen, zur Benennung, zur Bereitstellung verteilter Synchronisation und zur Bereitstellung von Gruppendiensten. Weitere Informationen zu ZooKeeper finden Sie unter <http://zookeeper.apache.org/>.

In der folgenden Tabelle sind die Version von ZooKeeper aufgeführt, die in der neuesten Version der Amazon EMR 7.x-Serie enthalten ist, zusammen mit den Komponenten, mit denen Amazon EMR installiert wird. ZooKeeper

Informationen zur Version der Komponenten, mit denen ZooKeeper in dieser Version installiert wurde, finden Sie unter Komponentenversionen von [Version 7.1.0](#).

ZooKeeper Versionsinformationen für emr-7.1.0

| Amazon-EMR-Version<br>sbezeichnung | ZooKeeper Version | Komponenten, die mit installiert wurden ZooKeeper                                                                                                                                                                                                                  |
|------------------------------------|-------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-7.1.0                          | ZooKeeper 3.9.1   | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |

In der folgenden Tabelle sind die Version von ZooKeeper aufgeführt, die in der neuesten Version der Amazon EMR 6.x-Serie enthalten ist, zusammen mit den Komponenten, mit denen Amazon EMR installiert wird. ZooKeeper

Informationen zur Version der Komponenten, mit denen ZooKeeper in dieser Version installiert wurde, finden Sie unter Komponentenversionen von [Version 6.15.0](#).

## ZooKeeper Versionsinformationen für emr-6.15.0

| Amazon-EMR-Version<br>sbezeichnung | ZooKeeper Version | Komponenten, die mit installiert wurden ZooKeeper                                                                                                                                                                                                                   |
|------------------------------------|-------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.15.0                         | ZooKeeper 3.5.10  | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |

In der folgenden Tabelle sind die Version von ZooKeeper aufgeführt, die in der neuesten Version der Amazon EMR 5.x-Serie enthalten ist, zusammen mit den Komponenten, mit denen Amazon EMR installiert wird. ZooKeeper

Informationen zur Version der Komponenten, mit denen ZooKeeper in dieser Version installiert wurde, finden Sie unter [Komponentenversionen von Version 5.36.2](#).

## ZooKeeper Versionsinformationen für emr-5.36.2

| Amazon-EMR-Version<br>sbezeichnung | ZooKeeper Version | Komponenten, die mit installiert wurden ZooKeeper                                                                                                                                                                                                                   |
|------------------------------------|-------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.36.2                         | ZooKeeper 3.4.14  | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |

## Themen

- [ZooKeeper Verlauf der Veröffentlichung](#)

## ZooKeeper Verlauf der Veröffentlichung

In der folgenden Tabelle sind die Versionen von aufgeführt, die in jeder Release-Version von Amazon EMR ZooKeeper enthalten sind, zusammen mit den Komponenten, die mit der Anwendung installiert wurden. Informationen zu den Komponentenversionen in den einzelnen Versionen finden Sie im Abschnitt Komponentenversion für Ihre Version in [Amazon-EMR-7.x-Versionen](#), [Amazon-EMR-6.x-Versionen](#) oder [Amazon-EMR-5.x-Versionen](#).

### ZooKeeper Versionsinformationen

| Amazon-EMR-Version<br>sbezeichnung | ZooKeeper Version | Komponenten, die installiert<br>wurden mit ZooKeeper                                                                                                                                                                                                               |
|------------------------------------|-------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.36.2                         | 3.4.14            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-7.1.0                          | 3.9.1             | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |



| Amazon-EMR-Version<br>sbezeichnung | ZooKeeper Version | Komponenten, die installiert<br>wurden mit ZooKeeper                                                                                                                                                                                                               |
|------------------------------------|-------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-7.0.0                          | 3.5.10            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-6.15.0                         | 3.5.10            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-6.14.0                         | 3.5.10            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |

| Amazon-EMR-Version<br>sbezeichnung | ZooKeeper Version | Komponenten, die installiert<br>wurden mit ZooKeeper                                                                                                                                                                                                              |
|------------------------------------|-------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.13.0                         | 3.5.10            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-6.12.0                         | 3.5.10            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-6.11.1                         | 3.5.10            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |

| Amazon-EMR-Version<br>sbezeichnung | ZooKeeper Version | Komponenten, die installiert<br>wurden mit ZooKeeper                                                                                                                                                                                                              |
|------------------------------------|-------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.11.0                         | 3.5.10            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-6.10.1                         | 3.5.10            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-6.10.0                         | 3.5.10            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |

| Amazon-EMR-Version<br>sbezeichnung | ZooKeeper Version | Komponenten, die installiert<br>wurden mit ZooKeeper                                                                                                                                                                                                              |
|------------------------------------|-------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.9.1                          | 3.5.10            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-6.9.0                          | 3.5.10            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-6.8.1                          | 3.5.10            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |

| Amazon-EMR-Version<br>sbezeichnung | ZooKeeper Version | Komponenten, die installiert<br>wurden mit ZooKeeper                                                                                                                                                                                                              |
|------------------------------------|-------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.8.0                          | 3.5.10            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-6.7.0                          | 3.5.7             | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-5.36.1                         | 3.4.14            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |

| Amazon-EMR-Version<br>sbezeichnung | ZooKeeper Version | Komponenten, die installiert<br>wurden mit ZooKeeper                                                                                                                                                                                                               |
|------------------------------------|-------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.36.0                         | 3.4.14            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-6.6.0                          | 3.5.7             | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-5.35.0                         | 3.4.14            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |

| Amazon-EMR-Version<br>sbezeichnung | ZooKeeper Version | Komponenten, die installiert<br>wurden mit ZooKeeper                                                                                                                                                                                                              |
|------------------------------------|-------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.5.0                          | 3.5.7             | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-6.4.0                          | 3.5.7             | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-6.3.1                          | 3.4.14            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |

| Amazon-EMR-Version<br>sbezeichnung | ZooKeeper Version | Komponenten, die installiert<br>wurden mit ZooKeeper                                                                                                                                                                                                               |
|------------------------------------|-------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.3.0                          | 3.4.14            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-6.2.1                          | 3.4.14            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-6.2.0                          | 3.4.14            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |



| Amazon-EMR-Version<br>sbezeichnung | ZooKeeper Version | Komponenten, die installiert<br>wurden mit ZooKeeper                                                                                                                                                                                                              |
|------------------------------------|-------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.1.1                          | 3.4.14            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-6.1.0                          | 3.4.14            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-6.0.1                          | 3.4.14            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |

| Amazon-EMR-Version<br>sbezeichnung | ZooKeeper Version | Komponenten, die installiert<br>wurden mit ZooKeeper                                                                                                                                                                                                               |
|------------------------------------|-------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.0.0                          | 3.4.14            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-5.34.0                         | 3.4.14            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-5.33.1                         | 3.4.14            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |

| Amazon-EMR-Version<br>sbezeichnung | ZooKeeper Version | Komponenten, die installiert<br>wurden mit ZooKeeper                                                                                                                                                                                                              |
|------------------------------------|-------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.33.0                         | 3.4.14            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-5.32.1                         | 3.4.14            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-5.32.0                         | 3.4.14            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |

| Amazon-EMR-Version<br>sbezeichnung | ZooKeeper Version | Komponenten, die installiert<br>wurden mit ZooKeeper                                                                                                                                                                                                              |
|------------------------------------|-------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.31.1                         | 3.4.14            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-5.31.0                         | 3.4.14            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-5.30.2                         | 3.4.14            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |

| Amazon-EMR-Version<br>sbezeichnung | ZooKeeper Version | Komponenten, die installiert<br>wurden mit ZooKeeper                                                                                                                                                                                                               |
|------------------------------------|-------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.30.1                         | 3.4.14            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-5.30.0                         | 3.4.14            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-5.29.0                         | 3.4.14            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |

| Amazon-EMR-Version<br>sbezeichnung | ZooKeeper Version | Komponenten, die installiert<br>wurden mit ZooKeeper                                                                                                                                                                                                               |
|------------------------------------|-------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.28.1                         | 3.4.14            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-5.28.0                         | 3.4.14            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-5.27.1                         | 3.4.14            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |

| Amazon-EMR-Version<br>sbezeichnung | ZooKeeper Version | Komponenten, die installiert<br>wurden mit ZooKeeper                                                                                                                                                                                                              |
|------------------------------------|-------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.27.0                         | 3.4.14            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-5.26.0                         | 3.4.14            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-5.25.0                         | 3.4.14            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |

| Amazon-EMR-Version<br>sbezeichnung | ZooKeeper Version | Komponenten, die installiert<br>wurden mit ZooKeeper                                                                                                                                                                                                              |
|------------------------------------|-------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.24.1                         | 3.4.13            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-5.24.0                         | 3.4.13            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-5.23.1                         | 3.4.13            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |



| Amazon-EMR-Version<br>sbezeichnung | ZooKeeper Version | Komponenten, die installiert<br>wurden mit ZooKeeper                                                                                                                                                                                                               |
|------------------------------------|-------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.23.0                         | 3.4.13            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-5.22.0                         | 3.4.13            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-5.21.2                         | 3.4.13            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |

| Amazon-EMR-Version<br>sbezeichnung | ZooKeeper Version | Komponenten, die installiert<br>wurden mit ZooKeeper                                                                                                                                                                                                               |
|------------------------------------|-------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.21.1                         | 3.4.13            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-5.21.0                         | 3.4.13            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-5.20.1                         | 3.4.13            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |

| Amazon-EMR-Version<br>sbezeichnung | ZooKeeper Version | Komponenten, die installiert<br>wurden mit ZooKeeper                                                                                                                                                                                                               |
|------------------------------------|-------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.20.0                         | 3.4.13            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-5.19.1                         | 3.4.13            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-5.19.0                         | 3.4.13            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |

| Amazon-EMR-Version<br>sbezeichnung | ZooKeeper Version | Komponenten, die installiert<br>wurden mit ZooKeeper                                                                                                                                                                                                               |
|------------------------------------|-------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.18.1                         | 3.4.12            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-5.18.0                         | 3.4.12            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-5.17.2                         | 3.4.12            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |

| Amazon-EMR-Version<br>sbezeichnung | ZooKeeper Version | Komponenten, die installiert<br>wurden mit ZooKeeper                                                                                                                                                                                                               |
|------------------------------------|-------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.17.1                         | 3.4.12            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-5.17.0                         | 3.4.12            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-5.16.1                         | 3.4.12            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |

| Amazon-EMR-Version<br>sbezeichnung | ZooKeeper Version | Komponenten, die installiert<br>wurden mit ZooKeeper                                                                                                                                                                                                              |
|------------------------------------|-------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.16.0                         | 3.4.12            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-5.15.1                         | 3.4.12            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-5.15.0                         | 3.4.12            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |

| Amazon-EMR-Version<br>sbezeichnung | ZooKeeper Version | Komponenten, die installiert<br>wurden mit ZooKeeper                                                                                                                                                                                                               |
|------------------------------------|-------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.14.2                         | 3.4.10            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-5.14.1                         | 3.4.10            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-5.14.0                         | 3.4.10            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |

| Amazon-EMR-Version<br>sbezeichnung | ZooKeeper Version | Komponenten, die installiert<br>wurden mit ZooKeeper                                                                                                                                                                                                               |
|------------------------------------|-------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.13.1                         | 3.4.10            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-5.13.0                         | 3.4.10            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-5.12.3                         | 3.4.10            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |



| Amazon-EMR-Version<br>sbezeichnung | ZooKeeper Version | Komponenten, die installiert<br>wurden mit ZooKeeper                                                                                                                                                                                                              |
|------------------------------------|-------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.12.2                         | 3.4.10            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-5.12.1                         | 3.4.10            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-5.12.0                         | 3.4.10            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |

| Amazon-EMR-Version<br>sbezeichnung | ZooKeeper Version | Komponenten, die installiert<br>wurden mit ZooKeeper                                                                                                                                                                                                               |
|------------------------------------|-------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.11.4                         | 3.4.10            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-5.11.3                         | 3.4.10            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-5.11.2                         | 3.4.10            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |

| Amazon-EMR-Version<br>sbezeichnung | ZooKeeper Version | Komponenten, die installiert<br>wurden mit ZooKeeper                                                                                                                                                                                                              |
|------------------------------------|-------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.11.1                         | 3.4.10            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-5.11.0                         | 3.4.10            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-5.10.1                         | 3.4.10            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |

| Amazon-EMR-Version<br>sbezeichnung | ZooKeeper Version | Komponenten, die installiert<br>wurden mit ZooKeeper                                                                                                                                                                                                              |
|------------------------------------|-------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.10.0                         | 3.4.10            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-5.9.1                          | 3.4.10            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-5.9.0                          | 3.4.10            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |

| Amazon-EMR-Version<br>sbezeichnung | ZooKeeper Version | Komponenten, die installiert<br>wurden mit ZooKeeper                                                                                                                                                                                                               |
|------------------------------------|-------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.8.3                          | 3.4.10            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-5.8.2                          | 3.4.10            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-5.8.1                          | 3.4.10            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |

| Amazon-EMR-Version<br>sbezeichnung | ZooKeeper Version | Komponenten, die installiert<br>wurden mit ZooKeeper                                                                                                                                                                                                               |
|------------------------------------|-------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.8.0                          | 3.4.10            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-5.7.1                          | 3.4.10            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-5.7.0                          | 3.4.10            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |

| Amazon-EMR-Version<br>sbezeichnung | ZooKeeper Version | Komponenten, die installiert<br>wurden mit ZooKeeper                                                                                                                                                                                                              |
|------------------------------------|-------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.6.1                          | 3.4.10            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-5.6.0                          | 3.4.10            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |
| emr-5.5.4                          | 3.4.10            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server                              |

| Amazon-EMR-Version<br>sbezeichnung | ZooKeeper Version | Komponenten, die installiert<br>wurden mit ZooKeeper                                                                                                                                                                                   |
|------------------------------------|-------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.5.3                          | 3.4.10            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server |
| emr-5.5.2                          | 3.4.10            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server |
| emr-5.5.1                          | 3.4.10            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server |



| Amazon-EMR-Version<br>sbezeichnung | ZooKeeper Version | Komponenten, die installiert<br>wurden mit ZooKeeper                                                                                                                                                                                   |
|------------------------------------|-------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.5.0                          | 3.4.10            | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server |
| emr-5.4.1                          | 3.4.9             | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server |
| emr-5.4.0                          | 3.4.9             | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server |

| Amazon-EMR-Version<br>sbezeichnung | ZooKeeper Version | Komponenten, die installiert<br>wurden mit ZooKeeper                                                                                                                                                                                   |
|------------------------------------|-------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.3.2                          | 3.4.9             | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server |
| emr-5.3.1                          | 3.4.9             | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server |
| emr-5.3.0                          | 3.4.9             | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server |

| Amazon-EMR-Version<br>sbezeichnung | ZooKeeper Version | Komponenten, die installiert<br>wurden mit ZooKeeper                                                                                                                                                                                   |
|------------------------------------|-------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.2.3                          | 3.4.9             | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server |
| emr-5.2.2                          | 3.4.9             | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server |
| emr-5.2.1                          | 3.4.9             | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server |

| Amazon-EMR-Version<br>sbezeichnung | ZooKeeper Version | Komponenten, die installiert<br>wurden mit ZooKeeper                                                                                                                                                                                   |
|------------------------------------|-------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.2.0                          | 3.4.8             | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server |
| emr-5.1.1                          | 3.4.8             | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server |
| emr-5.1.0                          | 3.4.8             | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server |

| Amazon-EMR-Version<br>sbezeichnung | ZooKeeper Version | Komponenten, die installiert<br>wurden mit ZooKeeper                                                                                                                                                                                   |
|------------------------------------|-------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.0.3                          | 3.4.8             | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server |
| emr-5.0.2                          | 3.4.8             | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server |
| emr-5.0.1                          | 3.4.8             | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server |

| Amazon-EMR-Version<br>sbezeichnung | ZooKeeper Version | Komponenten, die installiert<br>wurden mit ZooKeeper                                                                                                                                                                                   |
|------------------------------------|-------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.0.0                          | 3.4.8             | emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server |

# Konnektoren und Dienstprogramme

Amazon EMR bietet mehrere Konnektoren und Dienstprogramme für den Zugriff auf andere AWS Services als Datenquellen. Sie können in der Regel auf Daten in diesen Services innerhalb eines Programms zugreifen. Sie können beispielsweise einen Kinesis-Stream in einer Hive-Abfrage, einem Pig-Skript oder einer MapReduce Anwendung angeben und dann mit diesen Daten arbeiten.

## Themen

- [Tabellen in DynamoDB mit Amazon EMR exportieren, importieren, abfragen und verknüpfen](#)
- [Kinesis](#)
- [DistCp S3 \(s3-dist-cp\)](#)
- [Aufräumen nach fehlgeschlagenen DistCp S3-Jobs](#)

## Tabellen in DynamoDB mit Amazon EMR exportieren, importieren, abfragen und verknüpfen

### Note

Der Amazon EMR-DynamoDB-Connector ist als Open Source verfügbar auf GitHub [Weitere Informationen finden Sie unter \[https://github.com/aws-labs/. emr-dynamodb-connector\]\(https://github.com/aws-labs/emr-dynamodb-connector\)](#)

DynamoDB ist ein vollständig verwalteter NoSQL-Datenbankservice, der schnelle und vorhersehbare Leistung nahtlos skalierbar bereitstellt. Entwickler können eine Datenbanktabelle erstellen und den durch Anfragen erzeugten Datenverkehr oder den Speicher unbegrenzt erweitern. DynamoDB verteilt die Daten und den Datenverkehr für die Tabelle automatisch auf eine ausreichende Anzahl von Servern, um die vom Kunden angegebene Anforderungskapazität zu erreichen und die gespeicherte Datenmenge zu verarbeiten und dabei gleichzeitig eine konsistente, schnelle Leistung beizubehalten. Mit Amazon EMR und Hive können Sie große Datenmengen, z. B. in DynamoDB gespeicherte Date, schnell und effizient verarbeiten. Weitere Informationen über DynamoDB finden Sie im [Entwicklerhandbuch von Amazon DynamoDB](#).

Apache Hive ist ein Software-Layer, mit dem Sie MapReduce-Cluster mithilfe einer vereinfachten, SQL-ähnlichen Sprache namens HiveQL abfragen können. Die Software baut auf der Hadoop-

Architektur auf. Weitere Informationen zu Hive und HiveQL erhalten Sie in der [HiveQL-Sprachreferenz](#). Weitere Informationen über Hive und Amazon EMR erhalten Sie unter [Apache Hive](#).

Sie können Amazon EMR mit einer benutzerdefinierten Hive-Version verwenden, die Konnektivität mit DynamoDB umfasst, um Operationen für in DynamoDB gespeicherte Daten auszuführen:

- Laden von DynamoDB-Daten in das Hadoop Distributed File System (HDFS) und deren Verwendung als Input in einem Amazon-EMR-Cluster.
- Abfragen von Live DynamoDB-Daten mit SQL-ähnlichen Anweisungen (HiveQL).
- Verknüpfen von in DynamoDB gespeicherten Daten und deren Export oder Abfragen der verknüpften Daten.
- Exportieren von in DynamoDB gespeicherten Daten nach Amazon S3.
- Importieren von in Amazon S3 gespeicherten Daten in DynamoDB.

#### Note

Der Amazon-EMR-DynamoDB-Konnektor unterstützt keine Cluster, die zur Verwendung der [Kerberos-Authentifizierung](#) konfiguriert sind.

Für jede der folgenden Aufgaben starten Sie einen Amazon-EMR-Cluster, geben den Speicherort der Daten in DynamoDB an und führen Hive-Befehle zum Ändern der Daten in DynamoDB aus.

Es gibt mehrere Möglichkeiten, einen Amazon EMR-Cluster zu starten: Sie können die Amazon EMR-Konsole, die Befehlszeilenschnittstelle (CLI) verwenden oder Ihren Cluster mithilfe eines AWS SDK oder der Amazon EMR-API programmieren. Sie können auch bestimmen, ob ein Hive-Cluster interaktiv oder über ein Skript ausgeführt werden soll. In diesem Abschnitt wird beschrieben, wie Sie einen interaktive Hive-Cluster über die Amazon-EMR-Konsole und die CLI starten.

Die interaktive Verwendung von Hive ist eine hervorragende Möglichkeit zum Testen der Abfrageleistung und Optimieren Ihrer Anwendung. Nachdem Sie eine Gruppe von Hive-Befehlen eingerichtet haben, die regelmäßig ausgeführt werden, sollten Sie erwägen, ein Hive-Skript zu erstellen, das Amazon EMR für Sie ausführen kann.

#### Warning

Amazon-EMR-Lese- oder Schreibvorgänge in einer DynamoDB-Tabelle werden der festgelegten, bereitgestellten Durchsatzkapazität angerechnet, wodurch sich die Häufigkeit



von Ausnahmen des bereitgestellten Durchsatzes erhöht. Bei großen Anforderungen implementiert Amazon EMR Wiederholungen mit exponentiellem Backoff zur Verwaltung der Anforderungslast der DynamoDB-Tabelle. Das Ausführen von Amazon-EMR-Aufträgen gleichzeitig mit anderem Datenverkehr kann dazu führen, dass Sie die zugewiesene, bereitgestellte Durchsatzstufe überschreiten. Sie können dies überwachen, indem Sie die `ThrottleRequests`-Metrik in Amazon überprüfen CloudWatch. Wenn die Anforderungslast zu hoch ist, können Sie den Cluster neu starten und [Einstellung der Leserate in Prozent](#) oder [Einstellung der Schreibrate in Prozent](#) auf einen geringeren Wert festlegen, um die Amazon-EMR-Operationen zu drosseln. Weitere Informationen zu DynamoDB-Durchsatzeinstellungen finden Sie unter [Bereitgestellter Durchsatz](#).

Wenn eine Tabelle für den [On-Demand-Modus](#) konfiguriert ist, sollten Sie die Tabelle wieder in den Bereitstellungsmodus ändern, bevor Sie einen Export- oder Importvorgang ausführen. Pipelines benötigen ein Durchsatzverhältnis, um die zu verwendenden Ressourcen aus einer DynamoDB-Tabelle berechnen zu können. Im On-Demand-Modus wird der bereitgestellte Durchsatz entfernt. Um Durchsatzkapazität bereitzustellen, können Sie Amazon CloudWatch Events-Metriken verwenden, um den Gesamtdurchsatz auszuwerten, den eine Tabelle verwendet hat.

## Themen

- [Eine Hive-Tabelle einrichten, um Hive-Befehle auszuführen](#)
- [Hive-Befehlsbeispiele für das Exportieren, Importieren und Abfragen von Daten in DynamoDB](#)
- [Optimieren der Leistung von Amazon-EMR-Operationen in DynamoDB](#)

## Eine Hive-Tabelle einrichten, um Hive-Befehle auszuführen

Apache Hive ist eine Data Warehouse-Anwendung, mit der Sie Daten abfragen können, die in Amazon-EMR-Clustern enthalten sind. Dies erfolgt mit einer SQL-ähnlichen Sprache. Weitere Informationen zu Hive finden Sie unter <http://hive.apache.org/>.

Im folgenden Verfahren wird davon ausgegangen, dass Sie bereits einen Cluster und ein Amazon-EC2-Schlüsselpaar angegeben haben. Informationen zu den ersten Schritten beim Erstellen von Clustern finden Sie unter [Erste Schritte mit Amazon EMR](#) im Verwaltungshandbuch für Amazon EMR.

## Konfigurieren Sie Hive für die Verwendung MapReduce

Wenn Sie Hive in Amazon EMR zum Abfragen von DynamoDB-Tabellen verwenden, können Fehler auftreten, falls Hive die Standard-Engine Tez für die Ausführung nutzt. Aus diesem Grund empfehlen wir, bei der Erstellung eines Clusters mit Hive, der wie in diesem Abschnitt beschrieben in DynamoDB integriert ist, eine Konfigurationsklassifizierung zu verwenden, die Hive auf die Verwendung festlegt. MapReduce Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

Der folgende Ausschnitt zeigt die Konfigurationsklassifizierung und die Eigenschaft, die verwendet werden sollen, um sie MapReduce als Ausführungs-Engine für Hive festzulegen:

```
[
    {
        "Classification": "hive-site",
        "Properties": {
            "hive.execution.engine": "mr"
        }
    }
]
```

### Interaktives Ausführen von Hive-Befehlen

1. Herstellen einer Verbindung mit dem Master-Knoten. Weitere Informationen finden Sie unter [Mit SSH eine Verbindung zum Hauptknoten herstellen](#) im Verwaltungshandbuch für Amazon EMR.
2. Geben Sie an der Eingabeaufforderung für den aktuellen Master-Knoten hive ein.

Sie sollten eine Hive-Eingabeaufforderung sehen: hive>

3. Geben Sie einen Hive-Befehl ein, der die Daten in DynamoDB einer Tabelle in der Hive-Anwendung zuordnet. Diese Tabelle fungiert als Referenz für die in Amazon DynamoDB gespeicherten Daten. Die Daten werden nicht lokal in Hive gespeichert und alle Abfragen, die diese Tabelle verwenden, werden mit den Live-Daten in DynamoDB abgeglichen. Dabei wird die Lese- und Schreibkapazität der Tabelle jedes Mal verbraucht, wenn ein Befehl ausgeführt wird. Wenn Sie mehrere Hive-Befehle für denselben Datensatz ausführen möchten, sollten Sie ihn zunächst exportieren.

Das folgende Beispiel zeigt die Syntax für die Zuweisung einer Hive-Tabelle zu einer DynamoDB-Tabelle.

```
CREATE EXTERNAL TABLE hive_tablename
(hive_column1_name column1_datatype, hive_column2_name column2_datatype...)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "dynamodb_tablename",
"dynamodb.column.mapping" =
"hive_column1_name:dynamodb_attribute1_name, hive_column2_name:dynamodb_attribute2_name...")
```

Wenn Sie in Hive eine Tabelle mithilfe der DynamoDB erstellen möchten, müssen Sie sie unter Verwendung des Schlüsselworts EXTERNAL als externe Tabelle erstellen. Der Unterschied zwischen internen und externen Tabellen liegt darin, dass die Daten in internen Tabellen gelöscht werden, wenn die interne Tabelle gelöscht wird. Dies ist nicht das gewünschte Verhalten, wenn eine Verbindung zur Amazon DynamoDB besteht und somit werden nur externe Tabellen unterstützt.


Beispielsweise wird mit dem folgenden Befehl eine Tabelle mit dem Namen `hivetable1` in Hive erstellt, die auf die DynamoDB-Tabelle namens `dynamodhtable1` verweist. Die DynamoDB-Tabelle `dynamodhtable1` hat ein Primärschlüsselschema. `hash-and-range` Das Hash-Schlüsselement ist `name` (Zeichenfolgetyp), das Bereichsschlüsselement ist `year` (numerischer Typ) und jedes Element verfügt über einen Attributwert für `holidays` (Zeichenfolgetyp).

```
CREATE EXTERNAL TABLE hivetable1 (col1 string, col2 bigint, col3 array<string>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "dynamodhtable1",
"dynamodb.column.mapping" = "col1:name,col2:year,col3:holidays");
```

Zeile 1 verwendet die HiveQL-Anweisung `CREATE EXTERNAL TABLE`. Für `hivetable1` müssen Sie für jedes Attributnamen-Wertpaar in der DynamoDB-Tabelle eine Spalte erstellen und den Datentyp angeben. Für diese Werte muss die Groß- und Kleinschreibung nicht beachtet werden und Sie können den Spalten beliebige Namen geben (außer reservierten Wörtern).

Zeile 2 verwendet die `STORED BY`-Anweisung. Bei dem Wert `STORED BY` handelt es sich um den Namen der Klasse, die die Verbindung zwischen Hive und DynamoDB handhabt. Er sollte auf `'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'` festgelegt werden.

Zeile 3 verwendet die TBLPROPERTIES-Anweisung, um „hivetable1“ mit der richtigen Tabelle und dem Schema in DynamoDB zu verknüpfen. Stellen Sie TBLPROPERTIES die Werte für den `dynamodb.table.name`-Parameter und `dynamodb.column.mapping`-Parameter bereit. Bei diesen Werten muss die Groß- und Kleinschreibung beachtet werden.

 Note

Alle DynamoDB-Attributnamen für die Tabelle müssen über entsprechende Spalten in der Hive-Tabelle verfügen. Je nach Ihrer Amazon EMR-Version treten die folgenden Szenarien auf, wenn die one-to-one Zuordnung nicht existiert:

- In Amazon EMR Version 5.27.0 und höher verfügt der Connector über Validierungen, die eine one-to-one Zuordnung zwischen DynamoDB-Attributnamen und Spalten in der Hive-Tabelle sicherstellen. Wenn die Zuordnung nicht existiert, tritt ein Fehler auf. one-to-one
- Ab Amazon-EMR-Version 5.26.0 und früher enthält die Hive-Tabelle kein Name-Wert-Paar aus DynamoDB mehr. Wenn Sie die Primärschlüsselattribute der DynamoDB nicht zuordnen, generiert Hive einen Fehler. Wenn Sie ein nicht-primäres Schlüsselattribut nicht zuordnen, wird kein Fehler generiert, aber die Daten werden auch nicht in der Hive-Tabelle angezeigt. Wenn die Datentypen nicht übereinstimmen, ist der Wert Null.

Anschließend können Sie die Hive-Vorgänge in `hivetable1` starten. Abfragen für `hivetable1` werden intern über die DynamoDB-Tabelle `dynamodbtable1` Ihres DynamoDB-Kontos ausgeführt, wodurch bei jeder Ausführung Lese- oder Schreibeinheiten verbraucht werden.

Wenn Sie Hive-Abfragen über eine DynamoDB-Tabelle ausführen, müssen Sie sicherstellen, dass ausreichend Lesekapazitätseinheiten vorhanden sind.

Nehmen Sie zum Beispiel an, dass Sie 100 Lesekapazitätseinheiten für Ihre DynamoDB-Tabelle bereitgestellt haben. Dadurch können Sie 100 Lesevorgänge oder 409.600 Byte pro Sekunde ausführen. Wenn diese Tabelle 20 GB an Daten enthält (21.474.836.480 Bytes) und Ihre Hive-Abfrage einen vollständigen Tabellen-Scan durchführt, können Sie ermitteln, wie lange die Ausführung der Abfrage dauern wird:

$$21\,474\,836\,480 / 409\,600 = 52\,429 \text{ Sekunden} = 14,56 \text{ Stunden}$$

Die einzige Möglichkeit, um die erforderliche Zeit zu verkürzen, würde eine Änderung der Lesekapazitätseinheiten in der DynamoDB-Quelltabelle notwendig machen. Das Hinzufügen zusätzlicher Amazon-EMR-Knoten ist nicht hilfreich.

Der Fertigstellungsgrad in der Hive-Ausgabe wird dann aktualisiert, wenn ein oder mehrere Mapper-Prozesse abgeschlossen wurden. Für eine große DynamoDB-Tabelle mit geringer Lesekapazität wird der Fertigstellungsgrad der Ausgabe ggf. für längere Zeit nicht aktualisiert. Im oben genannten Fall wird der Auftrag mehrere Stunden lang als zu 0 % abgeschlossen angezeigt. Weitere Details zum Fortschritt Ihres Auftrags finden Sie in der Amazon-EMR-Konsole. Hier können Sie den Status einzelner Mapper-Aufgaben sowie Statistiken für Lesevorgänge anzeigen. Sie können sich auch bei der Hadoop-Schnittstelle auf dem Master-Knoten anmelden und die Hadoop-Statistiken anzeigen. Hier werden Ihnen der Status einzelner Map-Aufgaben sowie Statistiken für Lesevorgänge angezeigt. Weitere Informationen finden Sie unter den folgenden Themen:

- [Auf Hauptknoten gehostete Web-Schnittstellen](#)
- [Hadoop-Web-Schnittstellen anzeigen](#)

Weitere Informationen über HiveQL-Beispielanweisungen, um Aufgaben auszuführen, die bei der Ausführung von Aufgaben, wie dem Exportieren oder Importieren von DynamoDB-Daten sowie dem Verknüpfen von Tabellen behilflich sind, finden Sie unter [Hive-Befehlsbeispiele für das Exportieren, Importieren und Abfragen von Daten in DynamoDB](#).

### Abbrechen einer Hive-Anfrage

Wenn Sie eine Hive-Abfrage ausführen, enthält die Erstantwort vom Server den Befehl, die Anfrage abzubrechen. Um die Anfrage zu jedem Zeitpunkt während des Vorgangs abzubrechen, verwenden Sie den Kill Command (Beendigungsbefehl) aus der Serverantwort.

1. Geben Sie `Ctrl+C` ein, um den Befehlszeilen-Client zu beenden.
2. Geben Sie nach der Shell-Eingabeaufforderung den Kill Command (Beendigungsbefehl) aus der Erstantwort des Servers auf Ihre Anfrage ein.

Alternativ können Sie den folgenden Befehl aus der Befehlszeile des Master-Knotens ausführen, um den Hadoop-Auftrag zu beenden. Dabei ist die *Job-ID* die ID des Hadoop-Auftrags und kann von der Hadoop-Benutzeroberfläche aus abgerufen werden.

```
hadoop job -kill job-id
```

## Datentypen für Hive und DynamoDB

Die folgende Tabelle zeigt die verfügbaren Hive-Datentypen, den DynamoDB-Standardtyp, dem sie entsprechen, und die alternativen DynamoDB-Typen, denen sie auch zugeordnet werden können.

| Hive-Typ            | Standardtyp DynamoDB   | Alternative DynamoDB-Typ                                            |
|---------------------|------------------------|---------------------------------------------------------------------|
| Zeichenfolge        | String (S)             |                                                                     |
| Bigint oder Doppel  | Ganzzahl               |                                                                     |
| Binary              | Binary (B)             |                                                                     |
| boolesch            | boolescher Wert (BOOL) |                                                                     |
| Array               | list (L)               | Zahlensatz (ZS), ein Zeichenfolgesatz (ZFS) oder ein Binärsatz (BS) |
| map<string, string> | Element                | map (M)                                                             |
| map<string, ?>      | map (M)                |                                                                     |
|                     | null (NULL)            |                                                                     |

Wenn Sie Ihre Hive-Daten als entsprechenden alternativen DynamoDB-Typ schreiben möchten oder wenn Ihre DynamoDB-Daten Attributwerte eines alternativen -Typs enthalten, können Sie die Spalte und den DynamoDB-Typ mit dem Parameter `dynamodb.type.mapping` angeben. Das folgende Beispiel zeigt die Syntax für die Angabe einer alternativen Typzuweisung.

```
CREATE EXTERNAL TABLE hive_tablename (hive_column1_name column1_datatype,
  hive_column2_name column2_datatype...)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "dynamodb_tablename",
```

```
"dynamodb.column.mapping" =
  "hive_column1_name:dynamodb_attribute1_name,hive_column2_name:dynamodb_attribute2_name...",
"dynamodb.type.mapping" = "hive_column1_name:dynamodb_attribute1_datatype");
```

Der Parameter für die Typzuweisung ist optional und muss nur für die Spalten angegeben werden, die alternative Typen verwenden.

Beispielsweise erstellt der folgende Hive-Befehl eine Tabelle mit dem Namen `hivetable2`, die auf die DynamoDB-Tabelle `dynamodbtabelle2` verweist. Sie ist mit `hivetable1` vergleichbar, außer dass sie die `col3`-Spalte dem String Set(SS)-Typ zuordnet.

```
CREATE EXTERNAL TABLE hivetable2 (col1 string, col2 bigint, col3 array<string>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "dynamodbtabelle2",
"dynamodb.column.mapping" = "col1:name,col2:year,col3:holidays",
"dynamodb.type.mapping" = "col3:SS");
```

In Hive sind `hivetable1` und `hivetable2` identisch. Wenn jedoch Daten aus diesen Tabellen in die entsprechenden DynamoDB-Tabellen geschrieben werden, enthält `dynamodbtabelle1` Listen, während `dynamodbtabelle2` Zeichenfolgensätze enthält.

Wenn Sie Hive-null-Werte als Attribute des DynamoDB-null-Typs schreiben möchten, ist dies mit dem Parameter `dynamodb.null.serialization` möglich. Das folgende Beispiel zeigt die Syntax für die Angabe der null-Serialisierung.

```
CREATE EXTERNAL TABLE hive_tablename (hive_column1_name column1_datatype,
hive_column2_name column2_datatype...)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "dynamodb_tablename",
"dynamodb.column.mapping" =
  "hive_column1_name:dynamodb_attribute1_name,hive_column2_name:dynamodb_attribute2_name...",
"dynamodb.null.serialization" = "true");
```

Der Serialisierungsparameter `Null` ist optional und wird auf `false` gesetzt, wenn er nicht angegeben wird. Beachten Sie, dass DynamoDB `null`-Attribute unabhängig von der Parametereinstellung als `null`-Werte in Hive gelesen werden. Hive-Sammlungen mit `null`-Werten können nur dann in DynamoDB geschrieben werden, wenn der `Null-Serialisierungsparameter` als `true` angegeben ist. Andernfalls tritt ein Hive-Fehler auf.

Der `bigint`-Typ in Hive ist identisch mit dem Java-Typ "long" und der Hive-Doppeltyp entspricht mit dem Java-Doppeltyp im Hinblick auf die Genauigkeit. Das bedeutet, dass Sie für Ihre in DynamoDB gespeicherten numerischen Daten, deren Präzision höher liegt als in den Hive-Datentypen, bei der Verwendung von Hive für den Export oder Import oder bei einem Verweis auf die DynamoDB-Daten mit einem Präzisionsverlust oder einem Scheitern der Hive-Abfrage rechnen müssen.

Exporte des Binärtyps von der DynamoDB auf Amazon Simple Storage Service (Amazon S3) oder HDFS werden als Base64-kodierte Zeichenfolgen gespeichert. Wenn Sie Daten aus Amazon S3 oder HDFS in den DynamoDB-Binärtyp importieren, sollten die Daten als Base64-Zeichenfolge kodiert sein.

## Hive-Optionen

Sie können die folgenden Hive-Optionen festlegen, um die Datenübertragung aus Amazon DynamoDB zu verwalten. Diese Optionen gelten nur für die aktuelle Hive-Sitzung. Wenn Sie die Hive-Eingabeaufforderung schließen und später auf dem Cluster erneut öffnen, werden diese Einstellungen auf die Standardwerte zurückgesetzt.

| Hive-Optionen                                 | Beschreibung                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
|-----------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <code>dynamodb.throughput.read.percent</code> | <p>Legt die Rate der Lesevorgänge so fest, dass Ihre von DynamoDB bereitgestellte Durchsatzrate im für Ihre Tabelle zugewiesenen Bereich liegt. Der Wert liegt zwischen 0.1 und 1.5 (einschließlich).</p> <p>Der Wert für die Standardleserate liegt bei 0,5. Das bedeutet, dass Hive versuchen wird, die Hälfte der Lesekapazität, die für Ressourcen in der Tabelle bereitgestellt ist, zu verbrauchen. Eine Erhöhung dieses Werts auf über 0,5 verbessert die Leseanforderungsrate. Eine Verringerung dieses Werts unter 0,5 verringert die Leseanforderungsrate. Bei dieser Leserate handelt es sich um einen ungefähren Wert. Die tatsächliche Leserate variiert abhängig von Faktoren wie der Tatsache, ob eine einheitliche Verteilung der Schlüssel in DynamoDB vorliegt.</p> <p>Wenn Sie feststellen, dass der bereitgestellte Durchsatz häufig durch die Hive-Operation überschri</p> |



| Hive-Optionen | Beschreibung                                                                                                                                                                                                                                                                                                                                                                                                                                         |
|---------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|               | <p>Wenn die Live-Lesedatenverarbeitung zu langsam wird oder wenn der Live-Lesedatenverkehr zu sehr gedrosselt wird, reduzieren Sie diesen Wert auf unter 0.5. Wenn Sie über genügend Kapazität verfügen und sich schnellere Hive-Operationen wünschen, legen Sie den Wert auf über 0.5 fest. Sie können auch überzeichnen, indem Sie den Wert auf 1,5 festlegen, wenn Sie der Ansicht sind, dass ungenutzte Ein-/Ausgabevorgänge verfügbar sind.</p> |

| Hive-Optionen                                  | Beschreibung                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
|------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <code>dynamodb.throughput.write.percent</code> | <p>Legt die Rate der Schreibvorgänge so fest, dass Ihre von DynamoDB bereitgestellte Durchsatzrate im für Ihre Tabelle zugewiesenen Bereich liegt. Der Wert liegt zwischen 0.1 und 1.5 (einschließlich).</p> <p>Der Wert für die Standardschreibrate liegt bei 0,5. Das bedeutet, dass Hive versuchen wird, die Hälfte der Schreibkapazität, die für Ressourcen in der Tabelle bereitgestellt ist, zu verbrauchen. Eine Erhöhung dieses Werts auf über 0,5 verbessert die Schreibankforderungsrate. Eine Verringerung dieses Werts unter 0,5 verringert die Schreibankforderungsrate. Bei dieser Schreibrate handelt es sich um einen ungefähren Wert. Die tatsächliche Schreibrate variiert abhängig von Faktoren wie der Tatsache, ob eine einheitliche Verteilung der Schlüssel in DynamoDB vorliegt.</p> <p>Wenn Sie feststellen, dass der bereitgestellte Durchsatz häufig durch die Hive-Operation überschritten wird oder wenn der Live-Schreibdatenverkehr zu sehr gedrosselt wird, reduzieren Sie diesen Wert auf unter 0.5. Wenn Sie über genügend Kapazität verfügen und sich schnellere Hive-Operationen wünschen, legen Sie den Wert auf über 0.5 fest. Sie können auch überzeichnen, indem Sie den Wert auf 1,5 festlegen, wenn Sie der Ansicht sind, dass ungenutzte Ein-/Ausgabevorgänge verfügbar sind oder es sich hierbei um das anfängliche Hochladen von Daten in die Tabelle handelt und noch kein aktiver Datenverkehr vorliegt.</p> |
| <code>dynamodb.endpoint</code>                 | Geben Sie den Endpunkt für den DynamoDB-Service an. Weitere Informationen über die verfügbaren DynamoDB-Endpunkte finden Sie unter <a href="#">Regionen und Endpunkte</a> .                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |

| Hive-Optionen                        | Beschreibung                                                                                                                                                                                                                                            |
|--------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <code>dynamodb.max.map.tasks</code>  | Geben Sie die maximale Anzahl von Zuordnungsaufgaben beim Lesen von Daten aus DynamoDB an. Dieser Wert muss gleich oder größer 1 sein.                                                                                                                  |
| <code>dynamodb.retry.duration</code> | Geben Sie die Anzahl der Minuten an, die als Timeout-Dauer für das Wiederholen der Hive-Befehle verwendet werden soll. Bei diesem Wert muss es sich um eine Ganzzahl gleich oder größer als 0 handeln. Die Standard-Timeout-Dauer beträgt zwei Minuten. |

Diese Optionen werden über den Befehl SET, wie im folgenden Beispiel gezeigt, festgelegt.

```
SET dynamodb.throughput.read.percent=1.0;  
  
INSERT OVERWRITE TABLE s3_export SELECT *  
FROM hiveTableName;
```

## Hive-Befehlsbeispiele für das Exportieren, Importieren und Abfragen von Daten in DynamoDB

Die folgenden Beispiele verwenden Hive-Befehle für Operationen wie das Exportieren von Daten nach DynamoDB oder HDFS, das Importieren von Daten in Amazon S3, das Verknüpfen und Abfragen von Tabellen usw.

Operationen für eine Hive-Tabelle verweisen auf Daten, die in DynamoDB gespeichert sind. Hive-Befehle sind abhängig von den Einstellungen der DynamoDB-Tabelle für den bereitgestellten Durchsatz. Die Daten, die abgerufen werden, umfassen die Daten, die zum Zeitpunkt der Verarbeitung der Hive-Operationsanforderung durch DynamoDB in die DynamoDB-Tabelle geschrieben werden. Wenn der Datenabruf viel Zeit in Anspruch nimmt, wurden einige der vom Hive-Befehl zurückgegebenen Daten möglicherweise seit Starten des Hive-Befehls in DynamoDB aktualisiert.

Die Hive-Befehle `DROP TABLE` und `CREATE TABLE` gelten nur für die lokalen Tabellen in Hive und erstellen bzw. löschen keine Tabellen in DynamoDB. Wenn Ihre Hive-Abfrage auf eine Tabelle in DynamoDB verweist, muss diese bereits vorhanden sein, bevor Sie die Abfrage ausführen. Weitere Informationen zum Erstellen und Löschen von Tabellen in DynamoDB finden Sie unter [Arbeiten mit Tabellen in DynamoDB](#) im Amazon-DynamoDB-Entwicklerhandbuch.

#### Note

Wenn Sie eine Hive-Tabelle mit einem Speicherort in Amazon S3 verknüpfen, weisen Sie sie nicht dem Stammpfad des Buckets `s3://mybucket` zu, da dies Fehler verursachen kann, wenn Hive die Daten nach Amazon S3 schreibt. Weisen Sie die Tabelle stattdessen einem Unterpfad des Buckets zu ("`s3://mybucket/mypath`").

## Exportieren von Daten aus DynamoDB

Sie können mit Hive Daten aus DynamoDB exportieren.

So exportieren Sie eine DynamoDB-Tabelle in einen Amazon-S3-Bucket

- Erstellen Sie eine Hive-Tabelle, die auf in DynamoDB gespeicherte Daten verweist. Anschließend können Sie den Befehl `INSERT OVERWRITE` aufrufen, um die Daten in ein externes Verzeichnis zu schreiben. Im folgenden Beispiel ist `s3://bucketname/path/subpath/` ein gültiger Pfad in Amazon S3. Passen Sie die Spalten und Datentypen im `CREATE`-Befehl an, um mit den Werten in Ihrer DynamoDB übereinzustimmen. Sie können diese Methode zum Erstellen eines Archivs Ihrer DynamoDB-Daten in Amazon S3 verwenden.

```
CREATE EXTERNAL TABLE hiveTableName (col1 string, col2 bigint, col3 array<string>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "dynamodhtable1",
"dynamodb.column.mapping" = "col1:name,col2:year,col3:holidays");

INSERT OVERWRITE DIRECTORY 's3://bucketname/path/subpath/' SELECT *
FROM hiveTableName;
```

## So exportieren Sie eine DynamoDB-Tabelle in einen Amazon S3-Bucket mit Formatierung

- Erstellen Sie eine externe Tabelle, die auf einen Speicherort in Amazon S3 verweist. Dies wird unten als "s3\_export" gezeigt. Geben Sie während des CREATE-Aufrufs die Zeilenformatierung für die Tabelle an. Wenn Sie dann INSERT OVERWRITE für den Export von Daten aus DynamoDB in s3\_export verwenden, werden die Daten in dem angegebenen Format geschrieben. Im folgenden Beispiel werden die Daten als CSV-Werte geschrieben.

```
CREATE EXTERNAL TABLE hiveTableName (col1 string, col2 bigint, col3 array<string>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "dynamodatabase1",
"dynamodb.column.mapping" = "col1:name,col2:year,col3:holidays");

CREATE EXTERNAL TABLE s3_export(a_col string, b_col bigint, c_col array<string>)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
LOCATION 's3://bucketname/path/subpath/';

INSERT OVERWRITE TABLE s3_export SELECT *
FROM hiveTableName;
```

## So exportieren Sie eine DynamoDB-Tabelle in einen Amazon-S3-Bucket ohne Angabe einer Spaltenzuordnung

- Erstellen Sie eine Hive-Tabelle, die auf in DynamoDB gespeicherte Daten verweist. Dies ähnelt dem vorhergehenden Beispiel mit der Ausnahme, dass Sie keine Spaltenzuordnung angeben. Die Tabelle muss über genau einen Spaltentyp `map<string, string>` verfügen. Wenn Sie dann eine EXTERNAL-Tabelle in Amazon S3 erstellen, können Sie den Befehl INSERT OVERWRITE zum Schreiben der Daten von DynamoDB nach Amazon S3 aufrufen. Sie können diese Methode zum Erstellen eines Archivs Ihrer DynamoDB-Daten in Amazon S3 verwenden. Da es keine Spaltenzuordnung gibt, können Sie keine Tabellen abfragen, die auf diese Weise exportiert werden. Das Exportieren von Daten ohne Angabe einer Spaltenzuordnung ist ab Hive 0.8.1.5 verfügbar, was unter Amazon-EMR-AMI 2.2.x und höher unterstützt wird.

```
CREATE EXTERNAL TABLE hiveTableName (item map<string,string>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
```

```
TBLPROPERTIES ("dynamodb.table.name" = "dynamodhtable1");

CREATE EXTERNAL TABLE s3TableName (item map<string, string>)
ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' LINES TERMINATED BY '\n'
LOCATION 's3://bucketname/path/subpath/';

INSERT OVERWRITE TABLE s3TableName SELECT *
FROM hiveTableName;
```

So exportieren Sie eine DynamoDB-Tabelle in einen Amazon-S3-Bucket mit Datenkomprimierung

- Hive bietet mehrere Kompressions-Codecs an, die Sie während der Hive-Sitzung einrichten können. Dadurch werden die exportierten Daten im angegebenen Format komprimiert. Im folgenden Beispiel werden die exportierten Dateien mit dem Lempel-Ziv-Oberhumer (LZO)-Algorithmus komprimiert.

```
SET hive.exec.compress.output=true;
SET io.seqfile.compression.type=BLOCK;
SET mapred.output.compression.codec = com.hadoop.compression.lzo.LzopCodec;

CREATE EXTERNAL TABLE hiveTableName (col1 string, col2 bigint, col3 array<string>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "dynamodhtable1",
"dynamodb.column.mapping" = "col1:name,col2:year,col3:holidays");

CREATE EXTERNAL TABLE lzo_compression_table (line STRING)
ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' LINES TERMINATED BY '\n'
LOCATION 's3://bucketname/path/subpath/';

INSERT OVERWRITE TABLE lzo_compression_table SELECT *
FROM hiveTableName;
```

Die verfügbaren Kompressions-Codecs sind:

- [org.apache.hadoop.io.com/press](http://org.apache.hadoop.io.com/press). GzipCodec
- [org.apache.hadoop.io.com/press](http://org.apache.hadoop.io.com/press). DefaultCodec

- `com.hadoop.compression.lzo.LzoCodec`
- `com.hadoop.compression.lzo.LzopCodec`
- `org.apache.hadoop.io.compress.BZip2Codec`
- `org.apache.hadoop.io.com.compress.SnappyCodec`

So exportieren Sie eine DynamoDB-Tabelle nach HDFS

- Verwenden Sie den folgenden Hive-Befehl, wobei `hdfs://DirectoryName` ein gültiger HDFS-Pfad und eine Tabelle in Hive ist, die auf `hiveTableName` DynamoDB verweist. Dieser Exportvorgang ist schneller als der Export einer DynamoDB-Tabelle nach Amazon S3, da Hive 0.7.1.1 als Zwischenschritt beim Exportieren von Daten nach Amazon S3 HDFS verwendet. Das folgende Beispiel zeigt, wie Sie `dynamodb.throughput.read.percent` auf 1.0 festlegen, um die Leseanforderungsrate zu erhöhen.

```
CREATE EXTERNAL TABLE hiveTableName (col1 string, col2 bigint, col3 array<string>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "dynamodhtable1",
"dynamodb.column.mapping" = "col1:name,col2:year,col3:holidays");

SET dynamodb.throughput.read.percent=1.0;

INSERT OVERWRITE DIRECTORY 'hdfs:///directoryName' SELECT * FROM hiveTableName;
```

Sie können auch Daten nach HDFS via Formatierung und Kompression exportieren, wie oben beim Export nach Amazon S3 gezeigt. Ersetzen Sie dazu einfach das Amazon-S3-Verzeichnis in den Beispielen oben durch ein HDFS-Verzeichnis.

So lesen Sie nicht druckbare UTF-8-Zeichendaten in Hive

- Sie können nicht druckbare UTF-8-Zeichendaten mit Hive lesen und schreiben, indem Sie beim Erstellen der Tabelle die `STORED AS SEQUENCEFILE`-Klausel verwenden. A SequenceFile ist das Hadoop-Binärdateiformat. Sie müssen Hadoop verwenden, um diese Datei zu lesen. Das folgende Beispiel zeigt, wie Daten aus DynamoDB nach Amazon S3 exportiert werden. Sie können diese Funktionalität für die Verarbeitung von nicht druckbaren UTF-8-kodierten Zeichen verwenden.

```
CREATE EXTERNAL TABLE hiveTableName (col1 string, col2 bigint, col3 array<string>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "dynamodhtable1",
"dynamodb.column.mapping" = "col1:name,col2:year,col3:holidays");

CREATE EXTERNAL TABLE s3_export(a_col string, b_col bigint, c_col array<string>)
STORED AS SEQUENCEFILE
LOCATION 's3://bucketname/path/subpath/';

INSERT OVERWRITE TABLE s3_export SELECT *
FROM hiveTableName;
```

## Importieren von Daten in DynamoDB

Wenn Sie mit Hive Daten in DynamoDB schreiben möchten, stellen Sie sicher, dass die Anzahl der Schreibkapazitätseinheiten größer als die Anzahl der Mapper im Cluster ist. Zum Beispiel produzieren Cluster, die auf m1.xlarge-EC2-Instances ausgeführt werden, acht Mapper pro Instance. Bei einem Cluster mit 10 Instances ergäbe dies 80 Mapper. Wenn Ihre Schreibkapazitätseinheiten nicht größer als die Anzahl der Mapper im Cluster sind, verbraucht der Hive-Schreibvorgang möglicherweise den gesamten Schreibdurchsatz oder versucht, mehr Durchsatz zu verbrauchen, als zur Verfügung gestellt wird. Weitere Informationen zur Anzahl der Mapper, die von den einzelnen EC2-Instance-Typen produziert werden, finden Sie unter [Konfigurieren von Hadoop](#).

Die Anzahl der Mapper in Hadoop wird durch die Input Splits gesteuert. Wenn es zu wenig Splits gibt, ist Ihr Schreibbefehl ggf. nicht in der Lage, den gesamten verfügbaren Schreibdurchsatz zu verbrauchen.

Wenn ein Element mit demselben Schlüssel in der DynamoDB-Zieltabelle vorhanden ist, wird es überschrieben. Wenn kein Element mit dem Schlüssel in der DynamoDB-Zieltabelle vorhanden ist, wird das Element eingefügt.

So importieren Sie Daten aus Amazon S3 zu DynamoDB

- Sie können Amazon EMR (Amazon EMR) und Hive verwenden, um Daten von Amazon S3 nach DynamoDB zu schreiben.



```
CREATE EXTERNAL TABLE s3_import(a_col string, b_col bigint, c_col array<string>)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
LOCATION 's3://bucketname/path/subpath/';

CREATE EXTERNAL TABLE hiveTableName (col1 string, col2 bigint, col3 array<string>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "dynamodhtable1",
"dynamodb.column.mapping" = "col1:name,col2:year,col3:holidays");

INSERT OVERWRITE TABLE hiveTableName SELECT * FROM s3_import;
```

So importieren Sie eine Tabelle von einem Amazon-S3-Bucket in DynamoDB ohne Angabe einer Spaltenzuordnung

- Erstellen Sie eine EXTERNAL-Tabelle, die auf die in Amazon S3 gespeicherten Daten verweist, die zuvor von DynamoDB exportiert wurden. Stellen Sie vor dem Import sicher, dass die Tabelle in DynamoDB vorhanden ist und über das gleiche Schlüsselschema wie die zuvor exportierte DynamoDB-Tabelle verfügt. Die Tabelle muss außerdem über genau einen Spaltentyp `map<string, string>` verfügen. Wenn Sie dann eine Hive-Tabelle erstellen, die mit DynamoDB verknüpft ist, können Sie den Befehl `INSERT OVERWRITE` zum Schreiben der Daten von Amazon S3 nach DynamoDB aufrufen. Da es keine Spaltenzuordnung gibt, können Sie keine Tabellen abfragen, die auf diese Weise importiert werden. Das Importieren von Daten ohne Angabe einer Spaltenzuordnung ist ab Hive 0.8.1.5 verfügbar, was unter Amazon EMR AMI 2.2.3.x und höher unterstützt wird.

```
CREATE EXTERNAL TABLE s3TableName (item map<string, string>)
ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' LINES TERMINATED BY '\n'
LOCATION 's3://bucketname/path/subpath/';

CREATE EXTERNAL TABLE hiveTableName (item map<string,string>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "dynamodhtable1");

INSERT OVERWRITE TABLE hiveTableName SELECT *
FROM s3TableName;
```

## So importieren Sie eine Tabelle von HDFS in DynamoDB

- Sie können mit Amazon EMR und Hive Daten von HDFS nach DynamoDB schreiben.

```
CREATE EXTERNAL TABLE hdfs_import(a_col string, b_col bigint, c_col array<string>)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
LOCATION 'hdfs:///directoryName';

CREATE EXTERNAL TABLE hiveTableName (col1 string, col2 bigint, col3 array<string>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "dynamodbtable1",
"dynamodb.column.mapping" = "col1:name,col2:year,col3:holidays");

INSERT OVERWRITE TABLE hiveTableName SELECT * FROM hdfs_import;
```

## Abfragen von Daten in DynamoDB

Die folgenden Beispiele zeigen die verschiedenen Möglichkeiten zur Verwendung von Amazon EMR zum Abfragen von Daten in DynamoDB.

So finden Sie den größten Wert für eine zugeordnete Spalte (**max**)

- Verwenden Sie Hive-Befehle wie die folgenden. Im ersten Befehl wird durch die CREATE-Anweisung eine Hive-Tabelle erstellt, die auf in DynamoDB gespeicherte Daten verweist. Die SELECT-Anweisung verwendet dann diese Tabelle, um in DynamoDB gespeicherte Daten abzufragen. Im folgenden Beispiel wird nach der größten Bestellung eines bestimmten Kunden gesucht.

```
CREATE EXTERNAL TABLE hive_purchases(customerId bigint, total_cost double,
items_purchased array<String>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "Purchases",
"dynamodb.column.mapping" =
"customerId:CustomerId,total_cost:Cost,items_purchased:Items");
```

```
SELECT max(total_cost) from hive_purchases where customerId = 717;
```

So aggregieren Sie Daten mit der **GROUP BY**-Klausel

- Sie können die GROUP BY-Klausel zum Sammeln von Daten über mehrere Datensätze hinweg verwenden. Diese Klausel wird häufig in Verbindung mit einer Aggregationsfunktion wie „sum“, „count“, „min“ oder „max“ eingesetzt. Das folgende Beispiel gibt eine Liste der größten Bestellungen von Kunden zurück, die mehr als drei Bestellungen aufgegeben haben.

```
CREATE EXTERNAL TABLE hive_purchases(customerId bigint, total_cost double,  
items_purchased array<String>)  
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'  
TBLPROPERTIES ("dynamodb.table.name" = "Purchases",  
"dynamodb.column.mapping" =  
"customerId:CustomerId,total_cost:Cost,items_purchased:Items");  
  
SELECT customerId, max(total_cost) from hive_purchases GROUP BY customerId HAVING  
count(*) > 3;
```

So verknüpfen Sie zwei DynamoDB-Tabellen

- Im folgenden Beispiel werden zwei Hive-Tabellen Daten in DynamoDB zugeordnet. Dann wird ein Join zwischen diesen beiden Tabellen aufgerufen. Der Join wird auf dem Cluster verarbeitet und zurückgegeben. Der Join wird nicht in DynamoDB ausgeführt. In diesem Beispiel wird eine Liste von Kunden mit ihren Einkäufen angezeigt, die mehr als zwei Bestellungen getätigt haben.

```
CREATE EXTERNAL TABLE hive_purchases(customerId bigint, total_cost double,  
items_purchased array<String>)  
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'  
TBLPROPERTIES ("dynamodb.table.name" = "Purchases",  
"dynamodb.column.mapping" =  
"customerId:CustomerId,total_cost:Cost,items_purchased:Items");  
  
CREATE EXTERNAL TABLE hive_customers(customerId bigint, customerName string,  
customerAddress array<String>)  
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
```

```
TBLPROPERTIES ("dynamodb.table.name" = "Customers",
"dynamodb.column.mapping" =
"customerId:CustomerId,customerName:Name,customerAddress:Address");

Select c.customerId, c.customerName, count(*) as count from hive_customers c
JOIN hive_purchases p ON c.customerId=p.customerId
GROUP BY c.customerId, c.customerName HAVING count > 2;
```

So verknüpfen Sie zwei Tabellen aus verschiedenen Quellen

- Im folgenden Beispiel ist Customer\_S3 eine Hive-Tabelle, die eine CSV-Datei lädt, die in Amazon S3 gespeichert ist, und hive\_purchases ist eine Tabelle, die auf Daten in DynamoDB verweist. Das folgende Beispiel verbindet Kundendaten, die in Form einer CSV-Datei in Amazon S3 gespeichert sind, mit Daten in DynamoDB, um Daten für Bestellungen zurückzugeben, die von Kunden mit „Miller“ im Namen aufgegeben wurden.

```
CREATE EXTERNAL TABLE hive_purchases(customerId bigint, total_cost double,
items_purchased array<String>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "Purchases",
"dynamodb.column.mapping" =
"customerId:CustomerId,total_cost:Cost,items_purchased:Items");

CREATE EXTERNAL TABLE Customer_S3(customerId bigint, customerName string,
customerAddress array<String>)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
LOCATION 's3://bucketname/path/subpath/';

Select c.customerId, c.customerName, c.customerAddress from
Customer_S3 c
JOIN hive_purchases p
ON c.customerid=p.customerid
where c.customerName like '%Miller%';
```

**Note**

In den vorherigen Beispielen wurden aus Gründen der Klarheit und Vollständigkeit die CREATE TABLE-Anweisungen in jedes Beispiel eingefügt. Bei Ausführung mehrerer Abfragen oder Exportoperationen für eine bestimmte Hive-Tabelle müssen Sie die Tabelle nur einmal zu Beginn der Hive-Sitzung erstellen.

## Optimieren der Leistung von Amazon-EMR-Operationen in DynamoDB

Amazon-EMR-Operationen auf eine DynamoDB-Tabelle zählen als Lesevorgänge und unterliegen den für die Tabelle angegebenen Durchsatzeinstellungen. Amazon EMR implementiert eine eigene Logik, um zu versuchen, die Belastung Ihrer DynamoDB-Tabelle auszugleichen und so die Möglichkeit der Überschreitung des bereitgestellten Durchsatzes zu minimieren. Am Ende jeder Hive-Abfrage gibt Amazon EMR Informationen über den für die Verarbeitung der Abfrage verwendeten Cluster zurück, zum Beispiel darüber, wie oft der bereitgestellte Durchsatz überschritten wurde. Sie können diese Informationen sowie CloudWatch Metriken zu Ihrem DynamoDB-Durchsatz verwenden, um die Belastung Ihrer DynamoDB-Tabelle bei nachfolgenden Anfragen besser zu verwalten.

Die folgenden Faktoren beeinflussen die Hive-Abfrageleistung bei der Arbeit mit DynamoDB-Tabellen.

### Bereitgestellte Lesekapazitätseinheiten

Wenn Sie Hive-Abfragen über eine DynamoDB-Tabelle ausführen, müssen Sie sicherstellen, dass ausreichend Lesekapazitätseinheiten vorhanden sind.

Nehmen Sie zum Beispiel an, dass Sie 100 Lesekapazitätseinheiten für Ihre DynamoDB-Tabelle bereitgestellt haben. Dadurch können Sie 100 Lesevorgänge oder 409.600 Byte pro Sekunde ausführen. Wenn diese Tabelle 20 GB an Daten enthält (21.474.836.480 Bytes) und Ihre Hive-Abfrage einen vollständigen Tabellen-Scan durchführt, können Sie ermitteln, wie lange die Ausführung der Abfrage dauern wird:

$$21\,474\,836\,480 / 409\,600 = 52\,429 \text{ Sekunden} = 14,56 \text{ Stunden}$$

Die einzige Möglichkeit, um die erforderliche Zeit zu verkürzen, würde eine Änderung der Lesekapazitätseinheiten in der DynamoDB-Quellentabelle notwendig machen. Das Hinzufügen weiterer Knoten zum Amazon-EMR-Cluster würde nicht helfen.

Der Fertigstellungsgrad in der Hive-Ausgabe wird dann aktualisiert, wenn ein oder mehrere Mapper-Prozesse abgeschlossen wurden. Für eine große DynamoDB-Tabelle mit wenig bereitgestellter Lesekapazität wird die fertiggestellte Ausgabe in Prozent ggf. für längere Zeit nicht aktualisiert; im oben genannten Fall würde der Auftrag mehrere Stunden lang zu 0 % abgeschlossen angezeigt. Weitere Details zum Fortschritt Ihres Auftrags finden Sie in der Amazon-EMR-Konsole. Hier können Sie den Status einzelner Mapper-Aufgaben sowie Statistiken für Lesevorgänge anzeigen.

Sie können sich auch bei der Hadoop-Schnittstelle auf dem Master-Knoten anmelden und die Hadoop-Statistiken anzeigen. Hier werden Ihnen der Status einzelner Map-Aufgaben sowie Statistiken für Lesevorgänge angezeigt. Weitere Informationen finden Sie unter [Auf Hauptknoten gehostete Web-Schnittstellen](#) im Verwaltungshandbuch für Amazon EMR.

## Einstellung der Leserate in Prozent

Standardmäßig verwaltet Amazon EMR die Anforderungslast für Ihre DynamoDB-Tabelle entsprechend Ihres aktuell bereitgestellten Durchsatzes. Wenn Amazon EMR jedoch Informationen zu Ihrem Auftrag zurückgibt, die eine große Anzahl an Antworten enthalten, dass der bereitgestellte Durchsatz überschritten wurde, können Sie die Standardleserate bei Einrichtung der Hive-Tabelle mit dem Parameter `dynamodb.throughput.read.percent` anpassen. Weitere Informationen zum Festlegen des Parameters zur Leserate in Prozent finden Sie unter [Hive-Optionen](#).

## Einstellung der Schreibrate in Prozent

Standardmäßig verwaltet Amazon EMR die Anforderungslast für Ihre DynamoDB-Tabelle entsprechend Ihres aktuell bereitgestellten Durchsatzes. Wenn Amazon EMR jedoch Informationen zu Ihrem Auftrag zurückgibt, die eine große Anzahl an Antworten enthalten, dass der bereitgestellte Durchsatz überschritten wurde, können Sie die Standardschreibrate bei Einrichtung der Hive-Tabelle mit dem Parameter `dynamodb.throughput.write.percent` anpassen. Weitere Informationen zum Festlegen des Parameters zur Schreibrate in Prozent finden Sie unter [Hive-Optionen](#).

## Einstellung für das Wiederholungsintervall

Standardmäßig führt Amazon EMR eine Hive-Abfrage erneut aus, wenn nicht innerhalb von zwei Minuten (dem Standard-Wiederholungsintervall) ein Ergebnis zurückgegeben wurde. Sie können dieses Intervall anpassen, indem Sie bei der Ausführung einer Hive-Abfrage den Parameter `dynamodb.retry.duration` einstellen. Weitere Informationen zum Festlegen des Parameters zur Schreibrate in Prozent finden Sie unter [Hive-Optionen](#).

## Anzahl der Zuordnungs-Tasks

Die Zuordnungs-Daemons, die Hadoop startet, um Ihre Anforderungen zum Exportieren und Abfragen von in DynamoDB gespeicherten Daten zu verarbeiten, sind auf eine maximale Leserate von 1 MiB pro Sekunde begrenzt, um die verbrauchte Lesekapazität einzuschränken. Wenn zusätzlich bereitgestellter Durchsatz auf DynamoDB vorhanden ist, verbessern Sie die Performance von Hive-Export- und Abfrageoperationen, indem Sie die Anzahl der Zuordnungs-Daemons erhöhen. Zu diesem Zweck können Sie entweder die Anzahl der EC2-Instances in Ihrem Cluster oder die Anzahl der Zuordnungs-Daemons erhöhen, die auf jeder EC2-Instance ausgeführt werden.

Sie können die Anzahl der EC2-Instances in einem Cluster erhöhen, indem Sie den aktuellen Cluster beenden und mit einer größeren Anzahl von EC2-Instances erneut starten. Sie geben die Anzahl der EC2-Instances im Dialogfeld Configure EC2 Instances an, wenn Sie den Cluster von der Amazon-EMR-Konsole starten, oder mit der Option `--num-instances`, wenn Sie den Cluster von der Befehlszeilenschnittstelle starten.

Die Anzahl der Map-Tasks auf einer Instance hängt vom EC2-Instance-Typ ab. Weitere Informationen zu den unterstützten EC2 Instance-Typen und der Anzahl der von ihnen jeweils bereitgestellten Mapper finden Sie in der [Aufgabenkonfiguration](#). Hier finden Sie einen Abschnitt zur Aufgabenkonfiguration ("Task Configuration") für jede der unterstützten Konfigurationen.

Eine andere Möglichkeit, die Anzahl der Mapper-Daemons zu erhöhen, besteht darin, den Hadoop-Konfigurationsparameter `mapreduce.tasktracker.map.tasks.maximum` in einen höheren Wert zu ändern. Dies hat den Vorteil, dass Sie mehr Mapper erhalten, ohne die Anzahl oder die Größe der EC2-Instances erhöhen zu müssen, wodurch Sie Kosten sparen. Ein Nachteil ist, dass wenn Sie diesen Wert zu hoch einstellen, für die EC2-Instances in Ihrem Cluster nicht mehr genügend Arbeitsspeicher vorhanden ist. Zum Festlegen von `mapreduce.tasktracker.map.tasks.maximum` starten Sie den Cluster und geben einen Wert für `mapreduce.tasktracker.map.tasks.maximum` als Eigenschaft der Konfigurationsklassifizierung der `mapred-site` an. Dies wird im folgenden Beispiel veranschaulicht. Weitere Informationen finden Sie unter [Anwendungen konfigurieren](#).

```
{
  "configurations": [
    {
      "classification": "mapred-site",
      "properties": {
        "mapred.tasktracker.map.tasks.maximum": "10"
      }
    }
  ]
}
```

```
}  
  ]  
}
```

## Parallele Datenanforderungen

Mehrere Datenanforderungen, entweder von mehr als einem Benutzer oder von mehr als einer Anwendung, für eine einzelne Tabelle kann den bereitgestellten Lesedurchsatz erschöpfen und die Performance beeinträchtigen.

## Prozessdauer

Die Datenkonsistenz in DynamoDB hängt von der Reihenfolge der Lese- und Schreibvorgänge auf den einzelnen Knoten ab. Während eine Hive-Abfrage verarbeitet wird, kann eine andere Anwendung neue Daten in die DynamoDB-Tabelle laden oder vorhandene Daten ändern oder löschen. In diesem Fall enthalten die Ergebnisse der Hive-Abfrage möglicherweise nicht die Datenänderungen, die vorgenommen wurden, während die Abfrage ausgeführt wurde.

## Durchsatzüberschreitungen vermeiden

Wenn Sie Hive-Abfragen für DynamoDB ausführen, achten Sie darauf, dass Ihr bereitgestellter Durchsatz nicht überschritten wird, da sich dies nachteilig auf die Kapazität auswirkt, die für die DynamoDB: :Get-Aufrufe Ihrer Anwendung benötigt werden. Um sicherzustellen, dass dies nicht der Fall ist, sollten Sie das Lesevolumen und die Drosselung bei Anwendungsaufrufen regelmäßig überwachen, DynamoDB: :Get indem Sie die Protokolle und Messwerte in Amazon überprüfen.

CloudWatch

## Abfragezeit

Wenn Hive-Abfragen, die auf eine DynamoDB-Tabelle zugreifen, für Zeiten geplant werden, in denen wenig Anforderungen an die DynamoDB-Tabelle gerichtet werden, verbessert das die Leistung. Beispiel: Wenn die Mehrzahl der Anwendungsbenutzer in San Francisco lebt, können Sie die täglichen Daten um 04:00 Uhr PST, wenn die meisten Benutzer schlafen und keine Datensätze in der DynamoDB-Datenbank aktualisiert werden.

## Zeitbasierte Tabellen

Wenn die Daten als eine Reihe von zeitbasierten DynamoDB-Tabellen organisiert sind, z. B. eine Tabelle pro Tag, können Sie die Daten exportieren, wenn die Tabelle nicht mehr aktiv ist. Sie können auf diese Weise Daten fortlaufend nach Amazon S3 sichern.



## Archivierte Daten

Wenn Sie vorhaben, viele Hive-Abfragen für in DynamoDB gespeicherte Daten auszuführen und Ihre Anwendung archivierte Daten unterstützt, können Sie die Daten nach HDFS oder Amazon S3 exportieren und die Hive-Abfragen auf eine Kopie der Daten anstelle von DynamoDB ausführen. Dies spart Lesevorgänge und schont den bereitgestellten Durchsatz.

## Kinesis

Amazon EMR-Cluster können Amazon Kinesis Kinesis-Streams direkt lesen und verarbeiten, indem sie vertraute Tools aus dem Hadoop-Ökosystem wie Hive, Pig MapReduce, die Hadoop Streaming API und Cascading verwenden. Außerdem können Sie Echtzeitdaten aus Amazon Kinesis mit vorhandenen Daten auf Amazon S3, Amazon DynamoDB und HDFS in einem aktiven Cluster zusammenführen. Zur Nachbearbeitung können Sie die Daten aus Amazon EMR direkt in Amazon S3 oder DynamoDB laden. Informationen zu den wichtigsten Services und Preisen von Amazon Kinesis finden Sie auf der Seite zu [Amazon Kinesis](#).

## Wie kann ich die Integration von Amazon EMR und Amazon Kinesis verwenden?

Die Integration von Amazon EMR in Amazon Kinesis erleichtert bestimmte Szenarien, z. B.:

- Streaming-Protokollanalyse – Sie können Streaming-Webprotokolle analysieren, um eine Liste der 10 häufigsten Fehlertypen alle „x“ Minuten nach Region, Browser und Zugriffs-Domain zu erzeugen.
- Kundenbindung – Sie können Abfragen schreiben, die Clickstream-Daten aus Amazon Kinesis mit Informationen zu Werbekampagnen in einer DynamoDB-Tabelle verknüpfen, um die effektivsten Kategorien von Anzeigen zu bestimmen, die auf bestimmten Websites angezeigt werden.
- Interaktive Ad-hoc-Abfragen – Sie können Daten regelmäßig aus Amazon-Kinesis-Streams in HDFS laden und diese für schnelle, interaktive Analyseabfragen als lokale Impala-Tabelle zur Verfügung stellen.

## Analyse von Amazon-Kinesis-Streams mit Checkpointing

Benutzer können regelmäßige, auf Stapelverarbeitung basierende Analysen von Amazon-Kinesis-Streams in so genannten Iterationen ausführen. Da Amazon-Kinesis-Stream-Datensätze mithilfe von Sequenznummern abgerufen werden, erfolgt die Festlegung der Iterationsgrenzen anhand der Start-

und Endsequenznummern, die von Amazon EMR in einer DynamoDB-Tabelle gespeichert werden. Wenn z. B. `iteration0` endet, wird die letzte Sequenznummer in DynamoDB gespeichert, sodass bei Beginn des Auftrags `iteration1` nachfolgende Daten aus dem Stream abgerufen werden können. Diese Zuweisung von Iterationen in Stream-Daten wird als Checkpointing bezeichnet. Weitere Informationen finden Sie unter [Kinesis-Konnektor](#).

Wenn eine Iteration mit einem Checkpoint versehen wurde und der Auftrag eine Iteration nicht verarbeiten konnte, versucht Amazon EMR, die Datensätze in dieser Iteration erneut zu verarbeiten.

Checkpointing ist eine Funktion, mit der Sie Folgendes durchführen können:

- Starten von Datenverarbeitung nach einer Sequenznummer, die von einer vorherigen Abfrage verarbeitet wurde, die im selben Stream und mit selben logischen Namen ausgeführt wurde
- Erneutes Verarbeiten desselben Datenstapels aus Kinesis, der von einer früheren Abfrage verarbeitet wurde

Setzen Sie in Ihren Skripten den Parameter `kinesis.checkpoint.enabled` auf `true`, um Checkpointing zu aktivieren. Konfigurieren Sie außerdem die folgenden Parameter:

| Konfigurationseinstellung                                 | Beschreibung                                                                                                                                                                |
|-----------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <code>kinesis.checkpoint.metastore.table.name</code>      | Name der DynamoDB-Tabelle, in der Checkpoint-Informationen gespeichert werden                                                                                               |
| <code>kinesis.checkpoint.metastore.hash.key.name</code>   | Name des Hash-Schlüssels für die DynamoDB-Tabelle                                                                                                                           |
| <code>kinesis.checkpoint.metastore.hash.range.name</code> | Name des Bereichsschlüssels für die DynamoDB-Tabelle                                                                                                                        |
| <code>kinesis.checkpoint.logical.name</code>              | Ein logischer Name für die aktuelle Verarbeitung                                                                                                                            |
| <code>kinesis.checkpoint.iteration.no</code>              | Iterationsnummer für die Verarbeitung im Zusammenhang mit dem logischen Namen                                                                                               |
| <code>kinesis.rerun.iteration.without.wait</code>         | Boolescher Wert, der angibt, ob eine fehlgeschlagene Iteration erneut ausgeführt werden kann, ohne die Zeitbeschränkung abzuwarten. Der Standardwert ist <code>false</code> |

## Empfehlungen zu bereitgestellten IOPS für Amazon-DynamoDB-Tabellen

Der Amazon-EMR-Konnektor für Amazon Kinesis verwendet die DynamoDB-Datenbank als Sicherungsdatenbank für das Metadaten-Checkpointing. Sie müssen in DynamoDB eine Tabelle erstellen, bevor Sie Daten in einem Amazon-Kinesis-Stream mit einem Amazon-EMR-Cluster in Checkpoint-Intervallen verwenden können. Die Tabelle muss sich in derselben Region befinden wie Ihr Amazon-EMR-Cluster. Im Folgenden erhalten Sie allgemeine Empfehlungen für die Anzahl von IOPS, die Sie für Ihre DynamoDB-Tabellen bereitstellen sollten. Dabei sei  $j$  die maximale Anzahl der Hadoop-Aufträge (mit unterschiedlichen Kombinationen für logischer Name+Iterationsnummer), die gleichzeitig ausgeführt werden können, und  $s$  die maximale Anzahl der Shards, die ein Auftrag verarbeitet:

Für Read Capacity Units (Lese-Kapazitätseinheiten):  $j * s / 5$

Für Write Capacity Units (Schreibkapazitätseinheiten):  $j * s$

## Leistungsaspekte

Der Shard-Durchsatz für Amazon Kinesis ist direkt proportional zur Instance-Größe von Knoten in Amazon-EMR-Clustern und zur Datensatzgröße im Stream. Wir empfehlen die Verwendung von `m5.xlarge` oder größeren Instances für Haupt- und Core-Knoten.

## Die Amazon-Kinesis-Analyse mit Amazon EMR planen

Wenn Sie Daten in einem aktiven Amazon-Kinesis-Stream analysieren und dabei durch Timeouts und eine maximale Dauer der einzelnen Iterationen eingeschränkt sind, müssen Sie die Analyse unbedingt häufiger durchführen, um regelmäßig Details aus dem Stream zu erhalten. Es gibt verschiedene Möglichkeiten, solche Skripte und Abfragen in regelmäßigen Intervallen durchzuführen. Wir empfehlen für wiederkehrende Aufgaben wie diese [AWS Data Pipeline PigActivity](#) [AWS Data Pipeline HiveActivity](#) [AWS Data Pipeline](#). Weitere Informationen finden Sie unter [und im Developer Guide](#).

## Migration des Spark-Kinesis-Konnektors zu SDK 2.x für Amazon EMR 7.0

Das AWS SDK bietet eine Vielzahl von APIs und Bibliotheken für die Interaktion mit AWS Cloud-Computing-Diensten, z. B. die Verwaltung von Anmeldeinformationen und die Verbindung zu S3- und Kinesis-Diensten. Der Spark-Kinesis-Konnektor wird verwendet, um Daten aus Kinesis Data Streams zu verarbeiten, und die empfangenen Daten werden in der Ausführungs-Engine von Spark

transformiert und verarbeitet. Derzeit basiert dieser Konnektor auf 1.x von AWS SDK und Kinesis-client-library (KCL).

Im Rahmen der AWS SDK 2.x-Migration wird auch der Spark Kinesis-Konnektor entsprechend aktualisiert, sodass er mit dem SDK 2.x ausgeführt werden kann. In der Amazon-EMR-Version 7.0 enthält Spark das SDK-2.x-Upgrade, das in der Community-Version von Apache Spark noch nicht verfügbar ist. Wenn Sie den Spark-Kinesis-Konnektor aus einer Version unter 7.0 verwenden, müssen Sie Ihre Anwendungscodes für die Ausführung in SDK 2.x migrieren, bevor Sie zu Amazon EMR 7.0 migrieren können.

## Migrationshandbücher

In diesem Abschnitt werden die Schritte zur Migration einer Anwendung zum aktualisierten Spark-Kinesis-Konnektor beschrieben. Es enthält Anleitungen für die Migration zur Kinesis Client Library (KCL) 2.x, AWS Anmeldeinformationsanbieter und AWS Service-Clients in SDK 2.x. AWS Als Referenz enthält es auch ein [WordCount](#)Beispielprogramm, das den Kinesis-Konnektor verwendet.

### Themen

- [KCL von 1.x zu 2.x migrieren](#)
- [Anbieter von AWS Anmeldeinformationen von AWS SDK 1.x auf 2.x migrieren](#)
- [Migration von AWS Service-Clients von AWS SDK 1.x auf 2.x](#)
- [Codebeispiele für Streaming-Anwendungen](#)
- [Überlegungen zur Verwendung des aktualisierten Spark-Kinesis-Konnektors](#)

### KCL von 1.x zu 2.x migrieren

- Ebene und Dimensionen der Metriken in **KinesisInputDStream**

Wenn Sie einen `KinesisInputDStream` instanziiieren, können Sie die Metrikebene und die Dimensionen für den Stream steuern. Das folgende Beispiel zeigt, wie Sie diese Parameter mit KCL 1.x anpassen können:

```
import
  com.amazonaws.services.kinesis.clientlibrary.lib.worker.KinesisClientLibConfiguration
import com.amazonaws.services.kinesis.metrics.interfaces.MetricsLevel

val kinesisStream = KinesisInputDStream.builder
  .streamingContext(ssc)
  .streamName(streamName)
```

```

.endpointUrl(endpointUrl)
.regionName(regionName)
.initialPosition(new Latest())
.checkpointAppName(appName)
.checkpointInterval(kinesisCheckpointInterval)
.storageLevel(StorageLevel.MEMORY_AND_DISK_2)
.metricsLevel(MetricsLevel.DETAILED)

.metricsEnabledDimensions(KinesisClientLibConfiguration.DEFAULT_METRICS_ENABLED_DIMENSIONS.a
.build()

```

In KCL 2.x haben diese Konfigurationseinstellungen andere Paketnamen. Für die Migration zu 2.x:

1. Ändern Sie die Importanweisungen für `com.amazonaws.services.kinesis.clientlibrary.lib.worker.KinesisClientLibCon` und `com.amazonaws.services.kinesis.metrics.interfaces.MetricsLevel` zu `software.amazon.kinesis.metrics.MetricsLevel` bzw. `software.amazon.kinesis.metrics.MetricsUtil`.

```

// import com.amazonaws.services.kinesis.metrics.interfaces.MetricsLevel
import software.amazon.kinesis.metrics.MetricsLevel

// import
com.amazonaws.services.kinesis.clientlibrary.lib.worker.KinesisClientLibConfiguration
import software.amazon.kinesis.metrics.MetricsUtil

```

2. Ersetzen Sie die Zeile `metricsEnabledDimensionsKinesisClientLibConfiguration.DEFAULT_METRICS_ENABL` durch `metricsEnabledDimensionsSet(MetricsUtil.OPERATION_DIMENSION_NAME, MetricsUtil.SHARD_ID_DIMENSION_NAME)`

Im Folgenden finden Sie eine aktualisierte Version von `KinesisInputDStream` mit benutzerdefinierten Metrikebene und Metrikdimensionen:

```

import software.amazon.kinesis.metrics.MetricsLevel
import software.amazon.kinesis.metrics.MetricsUtil

val kinesisStream = KinesisInputDStream.builder
    .streamingContext(ssc)
    .streamName(streamName)

```

```

.endpointUrl(endpointUrl)
.regionName(regionName)
.initialPosition(new Latest())
.checkpointAppName(appName)
.checkpointInterval(kinesisCheckpointInterval)
.storageLevel(StorageLevel.MEMORY_AND_DISK_2)
.metricsLevel(MetricsLevel.DETAILED)
.metricsEnabledDimensions(Set(MetricsUtil.OPERATION_DIMENSION_NAME,
MetricsUtil.SHARD_ID_DIMENSION_NAME))
.build()

```

- Meldungshandler-Funktion in `KinesisInputDStream`

Bei der Instanziierung eines `KinesisInputDStream` können Sie auch eine „Meldungshandler-Funktion“ angeben, die einen Kinesis-Datensatz verwendet und ein generisches Objekt `T` zurückgibt, falls Sie andere in einem Datensatz enthaltene Daten wie den Partitionsschlüssel verwenden möchten.

In KCL 1.x lautet die Signatur der Meldungshandler-Funktion: `Record => T`, wobei `Record` für `com.amazonaws.services.kinesis.model.Record` steht. In KCL 2.x wurde die Signatur des Handlers in: `KinesisClientRecord => T`, where is geändert. `KinesisClientRecord` `software.amazon.kinesis.retrieval.KinesisClientRecord`

Es folgt ein Beispiel für die Bereitstellung eines Meldungshandlers in KCL 1.x:

```

import com.amazonaws.services.kinesis.model.Record

def addFive(r: Record): Int = JavaUtils.bytesToString(r.getData).toInt + 5
val stream = KinesisInputDStream.builder
  .streamingContext(ssc)
  .streamName(streamName)
  .endpointUrl(endpointUrl)
  .regionName(regionName)
  .initialPosition(new Latest())
  .checkpointAppName(appName)
  .checkpointInterval(Seconds(10))
  .storageLevel(StorageLevel.MEMORY_ONLY)
  .buildWithMessageHandler(addFive)

```

Für die Migration des Meldungshandlers:

1. Ändern Sie die Importanweisung für `com.amazonaws.services.kinesis.model.Record` zu `software.amazon.kinesis.retrieval.KinesisClientRecord`.

```
// import com.amazonaws.services.kinesis.model.Record
import software.amazon.kinesis.retrieval.KinesisClientRecord
```

2. Aktualisieren Sie die Methodensignatur des Meldungshandlers.

```
//def addFive(r: Record): Int = JavaUtils.bytesToString(r.getData).toInt + 5
def addFive = (r: KinesisClientRecord) => JavaUtils.bytesToString(r.data()).toInt
+ 5
```

Es folgt ein aktualisiertes Beispiel für die Bereitstellung eines Meldungshandlers in KCL 2.x:

```
import software.amazon.kinesis.retrieval.KinesisClientRecord

def addFive = (r: KinesisClientRecord) => JavaUtils.bytesToString(r.data()).toInt + 5
val stream = KinesisInputDStream.builder
  .streamingContext(ssc)
  .streamName(streamName)
  .endpointUrl(endpointUrl)
  .regionName(regionName)
  .initialPosition(new Latest())
  .checkpointAppName(appName)
  .checkpointInterval(Seconds(10))
  .storageLevel(StorageLevel.MEMORY_ONLY)
  .buildWithMessageHandler(addFive)
```

Weitere Informationen zur Migration von KCL 1.x zu 2.x finden Sie unter [Migrieren von Verbrauchern von KCL 1.x zu KCL 2.x](#).

Anbieter von AWS Anmeldeinformationen von AWS SDK 1.x auf 2.x migrieren

Anbieter von Anmeldeinformationen werden verwendet, um AWS Anmeldeinformationen für Interaktionen mit zu erhalten. AWS In SDK 2.x gibt es mehrere Schnittstellen- und Klassenänderungen im Zusammenhang mit den Anbietern von Anmeldeinformationen, die Sie [hier finden](#). Der Spark Kinesis-Konnektor hat eine Schnittstelle

(`org.apache.spark.streaming.kinesis.SparkAWSCredentials`) und Implementierungsklassen definiert, die die Version 1.x von AWS Credential Providern zurückgeben. Diese Anbieter von Anmeldeinformationen werden bei der Initialisierung von Kinesis-Clients benötigt. Wenn Sie die Methode beispielsweise `SparkAWSCredentials.provider` in den Anwendungen verwenden, müssten Sie die Codes aktualisieren, um die 2.x-Version der Credential Provider nutzen zu können. AWS

Im Folgenden finden Sie ein Beispiel für die Verwendung der Anmeldeinformationsanbieter in AWS SDK 1.x:

```
import org.apache.spark.streaming.kinesis.SparkAWSCredentials
import com.amazonaws.auth.AWSCredentialsProvider

val basicSparkCredentials = SparkAWSCredentials.builder
    .basicCredentials("accessKey", "secretKey")
    .build()

val credentialProvider = basicSparkCredentials.provider
assert(credentialProvider.isInstanceOf[AWSCredentialsProvider], "Type should be
    AWSCredentialsProvider")
```

Für die Migration zu 2.x:

1. Ändern Sie die Importanweisung für `com.amazonaws.auth.AWSCredentialsProvider` zu `software.amazon.awssdk.auth.credentials.AwsCredentialsProvider`

```
//import com.amazonaws.auth.AWSCredentialsProvider
import software.amazon.awssdk.auth.credentials.AwsCredentialsProvider
```

2. Aktualisieren Sie die verbleibenden Codes, die diese Klasse verwenden.

```
import org.apache.spark.streaming.kinesis.SparkAWSCredentials
import software.amazon.awssdk.auth.credentials.AwsCredentialsProvider

val basicSparkCredentials = SparkAWSCredentials.builder
    .basicCredentials("accessKey", "secretKey")
    .build()

val credentialProvider = basicSparkCredentials.provider
assert (credentialProvider.isInstanceOf[AwsCredentialsProvider], "Type should be
    AwsCredentialsProvider")
```



## Migration von AWS Service-Clients von AWS SDK 1.x auf 2.x

AWS Service-Clients haben in 2.x unterschiedliche Paketnamen (d. h. `software.amazon.awssdk`), wohingegen das SDK 1.x verwendet `com.amazonaws`. Weitere Informationen über die Änderungen in dieser Version finden Sie [hier](#). Wenn Sie diese Service-Clients in den Codes verwenden, müssten Sie die Clients entsprechend migrieren.

Im Folgenden finden Sie ein Beispiel für die Erstellung eines Clients in SDK 1.x:

```
import com.amazonaws.services.dynamodbv2.AmazonDynamoDBClient
import com.amazonaws.services.dynamodbv2.document.DynamoDB

AmazonDynamoDB ddbClient = AmazonDynamoDBClientBuilder.defaultClient();
AmazonDynamoDBClient ddbClient = new AmazonDynamoDBClient();
```

Für die Migration zu 2.x:

1. Ändern Sie die Importanweisungen für Service-Clients. Nehmen wir als Beispiel DynamoDB-Clients. Sie müssten `com.amazonaws.services.dynamodbv2.AmazonDynamoDBClient` oder `com.amazonaws.services.dynamodbv2.document.DynamoDB` zu `software.amazon.awssdk.services.dynamodb.DynamoDbClient` ändern.

```
// import com.amazonaws.services.dynamodbv2.AmazonDynamoDBClient
// import com.amazonaws.services.dynamodbv2.document.DynamoDB
import software.amazon.awssdk.services.dynamodb.DynamoDbClient
```

2. Die Codes aktualisieren, die die Clients initialisieren

```
// AmazonDynamoDB ddbClient = AmazonDynamoDBClientBuilder.defaultClient();
// AmazonDynamoDBClient ddbClient = new AmazonDynamoDBClient();

DynamoDbClient ddbClient = DynamoDbClient.create();
DynamoDbClient ddbClient = DynamoDbClient.builder().build();
```

Weitere Informationen zur Migration des AWS SDK von 1.x auf 2.x finden Sie unter [Was ist der Unterschied zwischen dem AWS SDK for Java 1.x und 2.x](#)

## Codebeispiele für Streaming-Anwendungen

```
import java.net.URI
```

```
import software.amazon.awssdk.auth.credentials.DefaultCredentialsProvider
import software.amazon.awssdk.http.apache.ApacheHttpClient
import software.amazon.awssdk.services.kinesis.KinesisClient
import software.amazon.awssdk.services.kinesis.model.DescribeStreamRequest
import software.amazon.awssdk.regions.Region
import software.amazon.kinesis.metrics.{MetricsLevel, MetricsUtil}

import org.apache.spark.SparkConf
import org.apache.spark.storage.StorageLevel
import org.apache.spark.streaming.{Milliseconds, StreamingContext}
import org.apache.spark.streaming.dstream.DStream.toPairDStreamFunctions
import org.apache.spark.streaming.kinesis.KinesisInitialPositions.Latest
import org.apache.spark.streaming.kinesis.KinesisInputDStream

object KinesisWordCountASLSDKV2 {

  def main(args: Array[String]): Unit = {
    val appName = "demo-app"
    val streamName = "demo-kinesis-test"
    val endpointUrl = "https://kinesis.us-west-2.amazonaws.com"
    val regionName = "us-west-2"

    // Determine the number of shards from the stream using the low-level Kinesis
    Client
    // from the AWS Java SDK.
    val credentialsProvider = DefaultCredentialsProvider.create
    require(credentialsProvider.resolveCredentials() != null,
      "No AWS credentials found. Please specify credentials using one of the methods
    specified " +
      "in https://docs.aws.amazon.com/sdk-for-java/latest/developer-guide/
    credentials.html")
    val kinesisClient = KinesisClient.builder()
      .credentialsProvider(credentialsProvider)
      .region(Region.US_WEST_2)
      .endpointOverride(URI.create(endpointUrl))
      .httpClientBuilder(ApacheHttpClient.builder())
      .build()
    val describeStreamRequest = DescribeStreamRequest.builder()
      .streamName(streamName)
      .build()
    val numShards = kinesisClient.describeStream(describeStreamRequest)
      .streamDescription
      .shards
  }
}
```

```
.size

// In this example, we are going to create 1 Kinesis Receiver/input DStream for
each shard.
// This is not a necessity; if there are less receivers/DStreams than the number of
shards,
// then the shards will be automatically distributed among the receivers and each
receiver
// will receive data from multiple shards.
val numStreams = numShards

// Spark Streaming batch interval
val batchInterval = Milliseconds(2000)

// Kinesis checkpoint interval is the interval at which the DynamoDB is updated
with information
// on sequence number of records that have been received. Same as batchInterval for
this
// example.
val kinesisCheckpointInterval = batchInterval

// Setup the SparkConfig and StreamingContext
val sparkConfig = new SparkConf().setAppName("KinesisWordCountASLSDKV2")
val ssc = new StreamingContext(sparkConfig, batchInterval)

// Create the Kinesis DStreams
val kinesisStreams = (0 until numStreams).map { i =>
  KinesisInputDStream.builder
    .streamingContext(ssc)
    .streamName(streamName)
    .endpointUrl(endpointUrl)
    .regionName(regionName)
    .initialPosition(new Latest())
    .checkpointAppName(appName)
    .checkpointInterval(kinesisCheckpointInterval)
    .storageLevel(StorageLevel.MEMORY_AND_DISK_2)
    .metricsLevel(MetricsLevel.DETAILED)
    .metricsEnabledDimensions(Set(MetricsUtil.OPERATION_DIMENSION_NAME,
MetricsUtil.SHARD_ID_DIMENSION_NAME))
    .build()
}

// Union all the streams
```

```
val unionStreams = ssc.union(kinesisStreams)

// Convert each line of Array[Byte] to String, and split into words
val words = unionStreams.flatMap(byteArray => new String(byteArray).split(" "))

// Map each word to a (word, 1) tuple so we can reduce by key to count the words
val wordCounts = words.map(word => (word, 1)).reduceByKey(_ + _)

// Print the first 10 wordCounts
wordCounts.print()

// Start the streaming context and await termination
ssc.start()
ssc.awaitTermination()
}
}
```

## Überlegungen zur Verwendung des aktualisierten Spark-Kinesis-Konnektors

- Wenn Ihre Anwendungen `Kinesis-producer-library` mit einer JDK-Version unter 11 verwenden, kann es zu Ausnahmen wie `java.lang.NoClassDefFoundError: javax/xml/bind/DatatypeConverter` kommen. Dies liegt daran, dass EMR 7.0 standardmäßig mit JDK 17 geliefert wird und J2EE-Module seit Java 11+ aus den Standardbibliotheken entfernt wurden. Dies könnte behoben werden, indem die folgende Abhängigkeit zur POM-Datei hinzugefügt wird. Ersetzen Sie die Bibliotheksversion nach Bedarf durch eine passende.

```
<dependency>
  <groupId>javax.xml.bind</groupId>
  <artifactId>jaxb-api</artifactId>
  <version>${jaxb-api.version}</version>
</dependency>
```

- Das Spark-Kinesis-Konnektor-JAR befindet sich nach der Erstellung eines EMR-Clusters unter diesem Pfad: `/usr/lib/spark/connector/lib/`

## DistCp S3 (s3-dist-cp)

Apache DistCp ist ein Open-Source-Tool, mit dem Sie große Datenmengen kopieren können. S3 DistCp ähnelt Amazon S3 DistCp, ist aber für die Verwendung mit AWS Amazon S3 optimiert. Der Befehl für S3 DistCp in Amazon EMR Version 4.0 und höher lautet `s3-dist-cp`, den Sie

als Schritt in einem Cluster oder in der Befehlszeile hinzufügen. Mit S3 DistCp können Sie große Datenmengen effizient von Amazon S3 nach HDFS kopieren, wo sie in nachfolgenden Schritten in Ihrem Amazon EMR-Cluster verarbeitet werden können. Sie können S3 auch verwenden DistCp , um Daten zwischen Amazon S3-Buckets oder von HDFS nach Amazon S3 zu kopieren. S3 DistCp ist skalierbarer und effizienter für das parallel Kopieren einer großen Anzahl von Objekten zwischen Buckets und AWS Konten.

Spezifische Befehle, die die Flexibilität von S3DistCP in realen Szenarien demonstrieren, finden Sie im Big Data-Blog unter [Sieben Tipps zur Verwendung von S3 DistCp](#). AWS

Wie bei DistCp S3 DistCp wird häufig MapReduce auf verteilte Weise kopiert. Kopier-, Fehlerbehandlungs-, Wiederherstellungs- und Berichterstellungsaufgaben werden auf mehrere Server verteilt. Weitere Informationen zum DistCp Open-Source-Projekt Apache finden Sie in der [DistCpAnleitung](#) in der Apache Hadoop-Dokumentation.

Wenn S3 DistCp einige oder alle der angegebenen Dateien nicht kopieren kann, schlägt der Cluster-Schritt fehl und gibt einen Fehlercode ungleich Null zurück. In diesem Fall bereinigt S3 teilweise kopierte Dateien DistCp nicht.



#### Important

S3 unterstützt DistCp keine Amazon S3 S3-Bucket-Namen, die den Unterstrich enthalten. S3 unterstützt DistCp keine Verkettung von Parquet-Dateien. Verwenden Sie stattdessen. PySpark Weitere Informationen finden Sie unter [Parquet-Dateien verketteten in Amazon EMR](#). Um Kopierfehler zu vermeiden, wenn Sie S3DistCP verwenden, um eine einzelne Datei (anstelle eines Verzeichnisses) von S3 nach HDFS zu kopieren, verwenden Sie Amazon-EMR-Version 5.33.0 oder höher oder Amazon-EMR-Version 6.3.0 oder höher.

## DistCp S3-Optionen

S3 ist zwar ähnlich DistCp, DistCp unterstützt aber verschiedene Optionen, mit denen Sie die Art und Weise ändern können, wie Daten kopiert und komprimiert werden.

Wenn Sie S3 aufrufen DistCp, können Sie die in der folgenden Tabelle beschriebenen Optionen angeben. Die Optionen werden dem Schritt mithilfe der Liste der Argumente hinzugefügt. Beispiele für die DistCp S3-Argumente sind in der folgenden Tabelle aufgeführt.

Option	Beschreibung	Erforderlich
<code>--src=LOCATION</code>	<p>Der Speicherort der zu kopierenden Daten. Hierbei kann es sich entweder um einen HDFS- oder Amazon-S3-Speicherort handeln.</p> <p>Beispiel: <code>--src=s3://DOC-EXAMPLE-BUCKET1/logs/j-3GYXXXXXX9I0J/node</code></p> <div style="border: 1px solid #f08080; border-radius: 10px; padding: 10px; margin-top: 10px;"> <p> <b>Important</b></p> <p>S3 unterstützt DistCp keine Amazon S3 S3-Bucket-Namen, die den Unterstrich enthalten.</p> </div>	Ja
<code>--dest=LOCATION</code>	<p>Das Ziel für die Daten. Hierbei kann es sich entweder um einen HDFS- oder Amazon-S3-Speicherort handeln.</p> <p>Beispiel: <code>--dest=hdfs:///output</code></p> <div style="border: 1px solid #f08080; border-radius: 10px; padding: 10px; margin-top: 10px;"> <p> <b>Important</b></p> <p>S3 unterstützt DistCp keine Amazon S3 S3-Bucket-Namen, die den Unterstrich enthalten.</p> </div>	Ja
<code>--srcPattern=PATTERN</code>	<p>Ein <a href="#">regulärer Ausdruck</a>, der die Kopieroperation auf eine Teilmenge der Daten unter <code>--src</code> filtert. Wenn weder <code>--srcPattern</code> noch <code>--groupBy</code> angegeben ist, werden alle Daten unter <code>--src</code> in <code>--dest</code> kopiert.</p> <p>Wenn das Argument des regulären Ausdrucks Sonderzeichen wie Sternchen (*) enthält, muss entweder der reguläre Ausdruck oder die gesamte</p>	Nein

Option	Beschreibung	Erforderlich
	<p>--args-Zeichenfolge in einfache Anführungszeichen (') gesetzt werden.</p> <p>Beispiel: --srcPattern=.*daemons.*-hadoop-.*</p>	
--groupBy=PATTERN	<p>Ein <a href="#">regulärer Ausdruck</a>, der S3 veranlasst, Dateien DistCp zu verketteten, die dem Ausdruck entsprechen. Sie können mit dieser Option beispielsweise alle Protokolldateien kombinieren, die innerhalb einer Stunde in einer einzigen Datei geschrieben wurden. Der verkettete Dateiname ist der Wert, der dem regulären Ausdruck für die Gruppe entspricht.</p> <p>Klammern geben an, wie Dateien gruppiert werden soll, wobei alle Elemente, die der eingeklammerten Anweisung entsprechen, zu einer einzigen Ausgabedatei kombiniert werden. Wenn der reguläre Ausdruck keine Anweisung in Klammern enthält, schlägt der Cluster beim Schritt S3 DistCp fehl und gibt einen Fehler zurück.</p> <p>Wenn das Argument des regulären Ausdrucks Sonderzeichen wie Sternchen (*) enthält, muss entweder der reguläre Ausdruck oder die gesamte --args-Zeichenfolge in einfache Anführungszeichen (') gesetzt werden.</p> <p>Wenn --groupBy angegeben ist, werden nur Dateien kopiert, die dem angegebenen Muster entsprechen. Sie müssen --groupBy und --srcPattern nicht gleichzeitig angeben.</p> <p>Beispiel: --groupBy=.*subnetid.*([0-9]+-[0-9]+-[0-9]+-[0-9]+).*</p>	Nein

Option	Beschreibung	Erforderlich
<code>--targetSize=SIZE</code>	<p>Die Größe (in Mebibyte (MiB)) der Dateien, die basierend auf der <code>--groupBy</code> -Option erstellt werden sollen. Dieser Wert muss eine Ganzzahl sein. Wenn dieser <code>--targetSize</code> Wert gesetzt ist, DistCp versucht S3, dieser Größe zu entsprechen. Die tatsächliche Größe der kopierten Dateien kann größer oder kleiner als dieser Wert sein. Aufträge werden basierend auf der Größe der Datendatei aggregiert. Deshalb ist es möglich, dass die Zieldateigröße der Quelldateigröße entspricht.</p> <p>Wenn die durch <code>--groupBy</code> verketteten Dateien größer als der Wert <code>--targetSize</code> sind, werden sie in Teildateien aufgelöst und sequenziell mit einem an das Ende angefügten numerischen Wert benannt. Eine Datei, die beispielsweise zu <code>myfile.gz</code> verkettet ist, wird wie folgt in mehrere Teile unterteilt: <code>myfile0.gz</code> , <code>myfile1.gz</code> usw.</p> <p>Beispiel: <code>--targetSize=2</code></p>	Nein
<code>--appendToLastFile</code>	<p>Gibt das Verhalten von S3 DistCp beim Kopieren von Dateien von Amazon S3 nach HDFS an, die bereits vorhanden sind. Es werden neue Dateien an vorhandene Dateien angefügt. Wenn Sie <code>--appendToLastFile</code> mit <code>--groupBy</code> verwenden, werden neue Daten an Dateien angefügt, die denselben Gruppen entsprechen. Diese Option respektiert auch das <code>--targetSize</code> -Verhalten, wenn sie mit <code>--groupBy</code> verwendet wird.</p>	Nein



Option	Beschreibung	Erforderlich
<code>--outputCodec=CODEC</code>	<p>Gibt den Komprimierungs-Codec für die kopierten Dateien an. Mögliche Werte sind: <code>gzip</code>, <code>gz</code>, <code>lzo</code>, <code>snappy</code> oder <code>none</code>. Sie können diese Option beispielsweise verwenden, um Eingabedateien, die mit Gzip komprimiert sind, in Ausgabedateien mit LZO-Komprimierung zu konvertieren oder um die Dateien im Rahmen der Kopieroperation zu dekomprimieren. Wenn Sie einen Ausgabe-Codec auswählen, wird dem Dateinamen die entsprechende Erweiterung angefügt (z. B. für <code>gz</code> und <code>gzip</code> die Erweiterung <code>.gz</code>). Wenn Sie keinen Wert für <code>--outputCodec</code> angeben, werden die Dateien ohne Änderung der Komprimierung kopiert.</p> <p>Beispiel: <code>--outputCodec=lzo</code></p>	Nein
<code>--s3ServerSideEncryption</code>	<p>Stellt sicher, dass die Zieldaten mit SSL übertragen und in Amazon S3 automatisch mit einem AWS serviceseitigen Schlüssel verschlüsselt werden. Beim Abrufen von Daten mit S3 DistCp werden die Objekte automatisch entschlüsselt. Wenn Sie versuchen, ein entschlüsseltes Objekt in einen Amazon-S3-Bucket zu kopieren, für den eine Verschlüsselung erforderlich ist, schlägt die Operation fehl. Weitere Informationen finden Sie unter <a href="#">Verwenden der Datenverschlüsselung</a>.</p> <p>Beispiel: <code>--s3ServerSideEncryption</code></p>	Nein

Option	Beschreibung	Erforderlich
<code>--deleteOnSuccess</code>	<p>Wenn der Kopiervorgang erfolgreich ist, veranlasst diese Option S3, die kopierten Dateien vom Quellspeicherort DistCp zu löschen. Dies ist nützlich, wenn Sie Ausgabedateien, wie Protokolldateien, von einem Speicherort an einen anderen als geplante Aufgabe kopieren und dieselben Dateien nicht zweimal kopiert werden sollen.</p> <p>Beispiel: <code>--deleteOnSuccess</code></p>	Nein
<code>--disableMultipartUpload</code>	<p>Deaktiviert die Verwendung des mehrteiligen Uploads.</p> <p>Beispiel: <code>--disableMultipartUpload</code></p>	Nein
<code>--multipartUploadChunkSize=SIZE</code>	<p>Die Größe jedes Teils in einem mehrteiligen Amazon-S3-Upload in MiB. S3 DistCp verwendet den mehrteiligen Upload, wenn Daten kopiert werden, die <code>multipartUploadChunkSize</code> größer sind als. Um die Arbeitsleistung zu verbessern, können Sie die Größe der einzelnen Teile erhöhen. Die Standardgröße beträgt 128 MiB.</p> <p>Beispiel: <code>--multipartUploadChunkSize=1000</code></p>	Nein
<code>--numberOfFiles</code>	<p>Stellt Ausgabedateien sequenzielle Nummern voran. Die Zählung beginnt bei 0, es sei denn, von <code>--startingIndex</code> wird ein anderer Wert angegeben.</p> <p>Beispiel: <code>--numberOfFiles</code></p>	Nein

Option	Beschreibung	Erforderlich
<code>--startingIndex=INDEX</code>	<p>Wird mit <code>--numberOfFiles</code> verwendet, um die erste Nummer in der Sequenz anzugeben.</p> <p>Beispiel: <code>--startingIndex=1</code></p>	Nein
<code>--outputManifest=FILENAME</code>	<p>Erstellt eine mit Gzip komprimierte Textdatei, die eine Liste aller von S3 kopierten Dateien enthält.</p> <p>DistCp</p> <p>Beispiel: <code>--outputManifest=manifest-1.gz</code></p>	Nein
<code>--previousManifest=PATH</code>	<p>Liest eine Manifestdatei, die bei einem früheren Aufruf von S3 DistCp mit dem <code>--outputManifest</code> Flag erstellt wurde. Wenn das <code>--previousManifest</code> Flag gesetzt ist, DistCp schließt S3 die im Manifest aufgelisteten Dateien vom Kopiervorgang aus. Wenn <code>--outputManifest</code> zusammen mit <code>--previousManifest</code> angegeben ist, erscheinen die im vorherigen Manifest aufgeführten Dateien auch in der neuen Manifestdatei, obwohl die Dateien nicht kopiert werden.</p> <p>Beispiel: <code>--previousManifest=/usr/bin/manifest-1.gz</code></p>	Nein
<code>--requirePreviousManifest</code>	<p>Erfordert ein vorheriges Manifest, das während eines vorherigen Aufrufs von S3 DistCp erstellt wurde. Wenn dieser Wert auf "false" festgelegt ist, wird kein Fehler generiert, wenn kein vorheriges Manifest angegeben ist. Der Standardwert ist „true“.</p>	Nein

Option	Beschreibung	Erforderlich
<code>--copyFromManifest</code>	<p>Kehrt das Verhalten <code>--previousManifest</code> um, dass S3 DistCp die angegebene Manifestdatei als Liste der zu kopierenden Dateien verwendet, anstatt als Liste von Dateien, die vom Kopieren ausgeschlossen werden sollen.</p> <p>Beispiel: <code>--copyFromManifest --previousManifest=/usr/bin/manifest-1.gz</code></p>	Nein
<code>--s3Endpoint=ENDPOINT</code>	<p>Gibt den Amazon-S3-Endpoint an, der beim Hochladen einer Datei verwendet werden soll. Mit dieser Option wird der Endpoint sowohl für die Quelle als auch das Ziel festgelegt. Falls die Option nicht festgelegt ist, wird als Standardendpoint <code>s3.amazonaws.com</code> verwendet. Eine Liste der Amazon-S3-Endpoints finden Sie unter <a href="#">Regionen und Endpunkte</a>.</p> <p>Beispiel: <code>--s3Endpoint=s3.eu-west-1.amazonaws.com</code></p>	Nein
<code>--storageClass=CLASS</code>	<p>Die zu verwendende Speicherklasse, wenn das Ziel Amazon S3 ist. Gültige Werte sind STANDARD und REDUCED_REDUNDANCY. Wenn diese Option nicht angegeben ist, DistCp versucht S3, die Speicherklasse beizubehalten.</p> <p>Beispiel: <code>--storageClass=STANDARD</code></p>	Nein

Option	Beschreibung	Erforderlich
<code>--srcPrefixesFile=PATH</code>	<p>Eine Textdatei in Amazon S3 (s3://), HDFS (hdfs://) oder dem lokalen Dateisystem (file:/) mit einer Liste von <code>src</code>-Präfixen, in der ein Präfix pro Zeile aufgeführt wird.</p> <p>Wenn <code>srcPrefixesFile</code> angegeben, listet S3 DistCp den <code>src</code>-Pfad nicht auf. Stattdessen wird eine Quellliste als kombiniertes Ergebnis der Auflistung aller in dieser Datei angegebenen Präfixe erstellt. Der relative Pfad im Vergleich zum <code>src</code>-Pfad, wird statt dieser Präfixe verwendet, um die Zielpfade zu generieren. Wenn <code>srcPattern</code> angegeben ist, wird dieser Wert auf die kombinierten Listenergebnisse der Quellpräfixe angewendet, um die Eingabe weiter zu filtern. Wenn <code>copyFromManifest</code> verwendet wird, werden die Objekte im Manifest kopiert und <code>srcPrefixesFile</code> wird ignoriert.</p> <p>Beispiel: <code>--srcPrefixesFile=PATH</code></p>	Nein

Zusätzlich zu den oben genannten Optionen DistCp implementiert S3 die [Tool-Schnittstelle](#), was bedeutet, dass es die generischen Optionen unterstützt.

## S3 DistCp als Schritt in einem Cluster hinzufügen

Sie können S3 aufrufen, DistCp indem Sie es als Schritt zu Ihrem Cluster hinzufügen. Sie können Schritte beim Start eines Clusters oder einem aktuell ausgeführten Cluster mithilfe der Konsole, Befehlszeilenschnittstelle oder API hinzufügen. Die folgenden Beispiele zeigen das Hinzufügen eines DistCp S3-Schritts zu einem laufenden Cluster. Weitere Informationen zum Hinzufügen von Schritten zu einem Cluster finden Sie unter [Aufträge an einen Cluster übermitteln](#) im Verwaltungshandbuch für Amazon EMR.

Um einem laufenden Cluster einen DistCp S3-Schritt hinzuzufügen, verwenden Sie den AWS CLI

Weitere Informationen zur Verwendung von Amazon EMR-Befehlen in der AWS CLI finden Sie in der [AWS CLI Befehlsreferenz](#).

- Um einem Cluster, der S3 aufruft, einen Schritt hinzuzufügen DistCp, übergeben Sie die Parameter, die angeben, wie S3 den Kopiervorgang ausführen DistCp soll, als Argumente.

Im folgenden Beispiel werden Daemon-Protokolle von Amazon S3 in `hdfs:///output` kopiert. Im folgenden Befehl:

- `--cluster-id` gibt den Cluster an.
- `Jar` ist der Speicherort der DistCp S3-JAR-Datei. Ein Beispiel für die Ausführung eines Befehls auf einem Cluster mithilfe von `command-runner.jar` finden Sie unter [Senden eines benutzerdefinierten JAR-Schritts zur Ausführung eines Skripts oder Befehls](#).
- `Args` ist eine durch Kommas getrennte Liste der Optionsname-Wert-Paare, die an S3 übergeben werden sollen. DistCp Eine vollständige Liste der verfügbaren Optionen finden Sie unter [DistCp S3-Optionen](#).

Um einem laufenden Cluster einen DistCp S3-Kopierschritt hinzuzufügen, fügen Sie wie *myStep.json* in diesem Beispiel Folgendes in eine JSON-Datei ein, die in Amazon S3 oder Ihrem lokalen Dateisystem gespeichert ist. Ersetzen Sie `j-3GYXXXXXX9I0K` durch Ihre Cluster-ID und *mybucket* durch Ihren Amazon-S3-Bucket-Namen.

```
[
  {
    "Name": "S3DistCp step",
    "Args": ["s3-dist-cp", "--s3Endpoint=s3.amazonaws.com", "--src=s3://mybucket/logs/j-3GYXXXXXX9I0J/node/", "--dest=hdfs:///output", "--srcPattern=.*[a-zA-Z,]+"],
    "ActionOnFailure": "CONTINUE",
    "Type": "CUSTOM_JAR",
    "Jar": "command-runner.jar"
  }
]
```

```
aws emr add-steps --cluster-id j-3GYXXXXXX9I0K --steps file:///./myStep.json
```

## Example Protokolldateien von Amazon S3 zu HDFS kopieren

Dieses Beispiel veranschaulicht, wie Protokolldateien in einem Amazon-S3-Bucket in HDFS kopiert werden, indem Sie einen Schritt zu einem aktuell ausgeführten Cluster hinzufügen. In diesem Beispiel werden die Daten, die in die Daemon-Protokolle kopiert werden, anhand der Option `--srcPattern` eingeschränkt.

Zum Kopieren von Protokolldateien von Amazon S3 in HDFS mithilfe der Option `--srcPattern` geben Sie Folgendes in einer JSON-Datei, die in Amazon S3 oder Ihrem lokalen Dateisystem gespeichert ist, als `myStep.json` in diesem Beispiel ein. Ersetzen Sie `j-3GYXXXXXX9I0K` durch Ihre Cluster-ID und `mybucket` durch Ihren Amazon-S3-Bucket-Namen.

```
[
  {
    "Name": "S3DistCp step",
    "Args": ["s3-dist-cp", "--s3Endpoint=s3.amazonaws.com", "--src=s3://mybucket/logs/j-3GYXXXXXX9I0J/node/", "--dest=hdfs:///output", "--srcPattern=.*daemons.*-hadoop-.*"],
    "ActionOnFailure": "CONTINUE",
    "Type": "CUSTOM_JAR",
    "Jar": "command-runner.jar"
  }
]
```

## Aufräumen nach fehlgeschlagenen DistCp S3-Jobs

Wenn S3 einige oder alle der angegebenen Dateien DistCp nicht kopieren kann, schlägt der Befehl oder der Clusterschritt fehl und gibt einen Fehlercode ungleich Null zurück. In diesem Fall bereinigt S3 teilweise kopierte Dateien DistCp nicht. Sie müssen sie manuell löschen.

Teilweise kopierte Dateien werden im tmp HDFS-Verzeichnis in Unterverzeichnissen mit der eindeutigen Kennung des DistCp S3-Jobs gespeichert. Die ID finden Sie in der Standardausgabe der Aufgabe.

Für einen DistCp S3-Job mit der ID können Sie `4b1c37bb-91af-4391-aa8-46a6067085a6` beispielsweise eine Verbindung zum Master-Knoten des Clusters herstellen und den folgenden Befehl ausführen, um die mit dem Job verknüpften Ausgabedateien anzuzeigen.

```
hdfs dfs -ls /tmp/4b1c37bb-91af-4391-aa8-46a6067085a6/output
```

Der Befehl gibt eine Liste von Dateien ähnlich der folgenden zurück:

Found 8 items

```
-rw-r--r-- 1 hadoop hadoop 0 2018-12-10 06:03 /tmp/4b1c37bb-91af-4391-  
aaf8-46a6067085a6/output/_SUCCESS  
-rw-r--r-- 1 hadoop hadoop 0 2018-12-10 06:02 /tmp/4b1c37bb-91af-4391-  
aaf8-46a6067085a6/output/part-r-00000  
-rw-r--r-- 1 hadoop hadoop 0 2018-12-10 06:02 /tmp/4b1c37bb-91af-4391-  
aaf8-46a6067085a6/output/part-r-00001  
-rw-r--r-- 1 hadoop hadoop 0 2018-12-10 06:02 /tmp/4b1c37bb-91af-4391-  
aaf8-46a6067085a6/output/part-r-00002  
-rw-r--r-- 1 hadoop hadoop 0 2018-12-10 06:03 /tmp/4b1c37bb-91af-4391-  
aaf8-46a6067085a6/output/part-r-00003  
-rw-r--r-- 1 hadoop hadoop 0 2018-12-10 06:03 /tmp/4b1c37bb-91af-4391-  
aaf8-46a6067085a6/output/part-r-00004  
-rw-r--r-- 1 hadoop hadoop 0 2018-12-10 06:03 /tmp/4b1c37bb-91af-4391-  
aaf8-46a6067085a6/output/part-r-00005  
-rw-r--r-- 1 hadoop hadoop 0 2018-12-10 06:03 /tmp/4b1c37bb-91af-4391-  
aaf8-46a6067085a6/output/part-r-00006
```

Sie können dann den folgenden Befehl ausführen, um das Verzeichnis und den gesamten Inhalt zu löschen.

```
hdfs dfs rm -rf /tmp/4b1c37bb-91af-4391-aaf8-46a6067085a6
```



# Befehle und Skripte auf einem Amazon-EMR-Cluster ausführen

In diesem Thema wird beschrieben, wie Sie einen Befehl oder ein Skript als Schritt auf Ihrem Cluster ausführen. Das Ausführen eines Befehls oder Skripts als Schritt ist eine der vielen Möglichkeiten, wie Sie [Arbeit an einen Cluster weiterleiten](#) können. Dies ist in den folgenden Situationen hilfreich:

- Wenn Sie keinen SSH-Zugriff auf Ihren Amazon-EMR-Cluster haben
- Wenn Sie einen Bash- oder Shell-Befehl zur Fehlerbehebung in Ihrem Cluster ausführen möchten

Sie können ein Skript ausführen, wenn Sie einen Cluster erstellen oder wenn sich Ihr Cluster im WAITING-Status befindet. Um ein Skript auszuführen, bevor die Schrittverarbeitung beginnt, verwenden Sie stattdessen eine Bootstrap-Aktion. Weitere Informationen finden Sie unter [Bootstrap-Aktionen zum Installieren zusätzlicher Software erstellen](#) im Verwaltungshandbuch für Amazon EMR.

Amazon EMR bietet die folgenden Tools, mit denen Sie Skripts, Befehle und andere Cluster-Programme ausführen können. Sie können beide Tools über die Amazon-EMR-Managementkonsole oder die AWS CLI aufrufen.

`command-runner.jar`

Befindet sich auf dem Amazon-EMR-AMI für Ihren Cluster. Sie können `command-runner.jar` verwenden, um Befehle auf Ihrem Cluster auszuführen. Sie geben den Pfad `command-runner.jar` an, ohne den vollständigen Pfad zu verwenden.

`script-runner.jar`

Gehostet auf Amazon S3 auf `s3://<region>.elasticmapreduce/libs/script-runner/script-runner.jar`, wobei `<region>` die Region ist, in der sich Ihr Amazon-EMR-Cluster befindet. Sie können `script-runner.jar` nutzen, um damit lokal oder auf Amazon S3 gespeicherte Skripte auf Ihrem Cluster auszuführen. Sie müssen die vollständige URI auf `script-runner.jar` angeben, wenn Sie einen Schritt einreichen.

## Reichen Sie einen benutzerdefinierten JAR-Schritt ein, um ein Skript oder einen Befehl auszuführen

Die folgenden AWS CLI Beispiele veranschaulichen einige gängige Anwendungsfälle von `command-runner.jar` und `script-runner.jar` auf Amazon EMR.

Example : Ausführen eines Befehls in einem Cluster mit **command-runner.jar**

Wenn Sie `command-runner.jar` verwenden, geben Sie Befehle, Optionen und Werte in der Argumentliste Ihres Schritts an.

Im folgenden AWS CLI-Beispiel wird ein Schritt an einen aktiven Cluster gesendet, der `command-runner.jar` auslöst. Der in der `Args`-Liste angegebene Befehl lädt ein Skript namens `my-script.sh` von Amazon S3 in das Hadoop-Benutzer-Home-Verzeichnis herunter. Der Befehl ändert dann die Berechtigungen des Skripts und führt `my-script.sh` aus.

Wenn Sie den AWS CLI verwenden, sollten die Elemente in Ihrer `Args` Liste durch Kommas getrennt sein, ohne Leerzeichen zwischen den Listenelementen. Verwenden Sie z. B. `Args=[example-command,example-option,"example option value"]` statt `Args=[example-command, example-option, "example option value"]`.

```
aws emr add-steps \  
--cluster-id j-2AXXXXXXGAPLF \  
--steps Type=CUSTOM_JAR,Name="Download a script from S3, change its permissions, and  
run it",ActionOnFailure=CONTINUE,Jar=command-runner.jar,Args=[bash,-c,"aws s3 cp s3://  
EXAMPLE-DOC-BUCKET/my-script.sh /home/hadoop; chmod u+x /home/hadoop/my-script.sh; cd /  
home/hadoop; ./my-script.sh"]
```

Example : Ausführen eines Skripts in einem Cluster mit **script-runner.jar**

Wenn Sie `script-runner.jar` verwenden, geben Sie das Skript, das Sie ausführen möchten, in der Argumentliste Ihres Schritts an.

Im folgenden AWS CLI-Beispiel wird ein Schritt an einen aktiven Cluster gesendet, der `script-runner.jar` auslöst. In diesem Fall wird das Skript mit dem Namen `my-script.sh` auf Amazon S3 gespeichert. Sie können auch lokale Skripte angeben, die auf dem Hauptknoten Ihres Clusters gespeichert sind.

```
aws emr add-steps \  
--cluster-id j-2AXXXXXXGAPLF \  

```

```
--steps Type=CUSTOM_JAR,Name="Run a script from S3 with script-
runner.jar",ActionOnFailure=CONTINUE,Jar=s3://us-west-2.elasticmapreduce/libs/script-
runner/script-runner.jar,Args=[s3://EXAMPLE-DOC-BUCKET/my-script.sh]
```

## Andere Verwendungsmöglichkeiten von **command-runner.jar**

Sie können `command-runner.jar` auch verwenden, um Arbeiten mit Tools wie `spark-submit` oder `hadoop-streaming` in einen Cluster einzureichen. Wenn Sie eine Anwendung mit `command-runner.jar` starten, geben Sie `CUSTOM_JAR` als Schritttyp an, anstatt einen Wert wie `SPARK`, `STREAMING` oder `PIG` zu verwenden. Die Verfügbarkeit der Tools hängt davon ab, welche Anwendungen Sie auf dem Cluster installiert haben.

Der folgende Beispielbefehl `command-runner.jar` dient zum Senden eines Schritts mit `spark-submit`. In der `Args`-Liste wird `spark-submit` als Befehl angegeben, gefolgt von der Amazon-S3-URI der Spark-Anwendung `my-app.py` mit Argumenten und Werten.

```
aws emr add-steps \
--cluster-id j-2AXXXXXXGAPLF \
--steps Type=CUSTOM_JAR,Name="Run spark-submit using command-
runner.jar",ActionOnFailure=CONTINUE,Jar=command-runner.jar,Args=[spark-submit,S3://
DOC-EXAMPLE-BUCKET/my-app.py,ArgName1,ArgValue1,ArgName2,ArgValue2]
```

In der folgenden Tabelle sind zusätzliche Tools aufgeführt, die Sie mit `command-runner.jar` verwenden können.

Name des Tools	Beschreibung
<code>hadoop-streaming</code>	Sendet ein Hadoop-Streaming-Programm. In der Konsole und einige SDKs ist dies ein Streaming-Schritt.
<code>hive-script</code>	Führt ein Hive-Skript aus. In der Konsole und in SDKs ist dies ein Hive-Schritt.
<code>pig-script</code>	Führt ein Pig-Skript aus. In der Konsole und in SDKs ist dies ein Pig-Schritt.
<code>spark-submit</code>	Führt eine Spark-Anwendung aus. In der Konsole ist dies ein Spark-Schritt.

Name des Tools	Beschreibung
hadoop-lzo	Führt die <a href="#">Hadoop-LZO-Indexerstellung</a> in einem Verzeichnis aus.
s3-dist-cp	Verteiltes Kopieren großer Datenmengen aus Amazon S3 in HDFS. Weitere Informationen finden Sie unter <a href="#">DistCp S3 (s3-dist-cp)</a> .

# AWS-Glossar

Die neueste AWS-Terminologie finden Sie im [AWS-Glossar](#) in der AWS-Glossar-Referenz.

Die vorliegende Übersetzung wurde maschinell erstellt. Im Falle eines Konflikts oder eines Widerspruchs zwischen dieser übersetzten Fassung und der englischen Fassung (einschließlich infolge von Verzögerungen bei der Übersetzung) ist die englische Fassung maßgeblich.