



Handbuch „Erste Schritte“

# Amazon Redshift



# Amazon Redshift: Handbuch „Erste Schritte“

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, auf eine Art und Weise, dass Kunden irreführt werden könnten oder Amazon schlecht gemacht oder diskreditiert werden könnte. Alle anderen Handelsmarken, die nicht im Besitz von Amazon sind, gehören den jeweiligen Besitzern, die möglicherweise zu Amazon gehören oder nicht, mit Amazon verbunden sind oder von Amazon gesponsert werden.

---

# Table of Contents

Amazon Redshift Serverless .....	1
Melde dich an für AWS .....	1
Erstellen eines Data Warehouse mit Amazon Redshift Serverless .....	1
Laden von Beispieldaten .....	4
Ausführen von Beispielabfragen .....	8
Laden von Daten aus Amazon S3 .....	10
Von Amazon Redshift bereitgestellte Cluster .....	18
Melden Sie sich an für AWS .....	20
Festlegen von Firewall-Regeln .....	21
Schritt 1: Einen Beispiel-Cluster erstellen .....	21
Schritt 2: Regeln für eingehenden Datenverkehr für SQL-Clients konfigurieren .....	25
Schritt 3: Gewähren Sie Zugriff auf einen SQL-Client und führen Sie Abfragen aus .....	26
Gewähren des Zugriffs auf Abfrage-Editor v2 .....	26
Schritt 4: Daten aus Amazon S3 in Amazon Redshift laden .....	27
Laden von Daten aus Amazon S3 mithilfe von SQL-Befehlen .....	28
Laden von Daten aus Amazon S3 mit dem Abfrage-Editor v2 .....	30
Erstellen Sie TICKIT-Daten in Ihrem Cluster .....	30
Schritt 5: Beispielabfragen mit dem Abfrage-Editor testen .....	31
Schritt 6: Umgebung zurücksetzen .....	32
Häufige Datenbankaufgaben .....	33
Verbinden mit Amazon Redshift .....	33
Schritt 1: Erstellen einer Datenbank .....	35
Schritt 2: Erstellen Sie einen Benutzer .....	35
Schritt 3: Erstellen Sie ein Schema .....	36
Schritt 4: Erstellen Sie eine Tabelle .....	38
Einfügen von Datenzeilen in eine Tabelle .....	39
Auswahl von Daten aus einer Tabelle .....	39
Schritt 5: Daten laden .....	40
Schritt 6: Fragen Sie die Systemtabellen ab .....	40
Anzeigen einer Liste von Tabellennamen .....	40
Anzeigen von Benutzern .....	42
Anzeigen aktueller Abfragen .....	43
Ermitteln Sie die Sitzungs-ID einer laufenden Abfrage .....	43
Schritt 7: Stornieren Sie eine Abfrage .....	44

---

Abbrechen einer Abfrage mit der Superuser-Warteschlange .....	46
Abfragen von externen Datenquellen .....	48
Abfragen von Data Lakes .....	48
Abfragen von Remote-Datenquellen .....	49
Zugriff auf Daten in anderen Clustern .....	49
Training von ML-Modellen mit Redshift-Daten .....	50
Amazon Redshift – Konzeptionelle Übersicht .....	51
Weitere Ressourcen .....	55
Dokumentverlauf .....	57
.....	lix

# Amazon Redshift Serverless

Wenn Sie Amazon Redshift Serverless zum ersten Mal verwenden, empfehlen wir Ihnen, die folgenden Abschnitte zu lesen, um Ihnen den Einstieg in die Verwendung von Amazon Redshift Serverless zu erleichtern. Der grundlegende Ablauf von Amazon Redshift Serverless besteht darin, Serverless-Ressourcen zu erstellen, eine Verbindung zu Amazon Redshift Serverless herzustellen, Beispieldaten zu laden und dann Abfragen für die Daten auszuführen. Bei Verwendung dieses Handbuchs haben Sie die Möglichkeit, Beispieldaten aus Amazon Redshift Serverless oder aus einem Amazon-S3-Bucket zu laden.

- [the section called “Melde dich an für AWS”](#)
- [the section called “Erstellen eines Data Warehouse mit Amazon Redshift Serverless”](#)
- [the section called “Laden von Daten aus Amazon S3”](#)

## Melde dich an für AWS

Wenn Sie noch kein AWS Konto haben, registrieren Sie sich für eines. Wenn Sie bereits ein Konto besitzen, können Sie diesen Schritt überspringen und Ihr vorhandenes Konto verwenden.

1. Öffnen Sie <https://portal.aws.amazon.com/billing/signup>.
2. Folgen Sie den Online-Anweisungen.

Wenn Sie sich für ein AWS Konto registrieren, wird ein Root-Benutzer für das AWS Konto erstellt. Der Root-Benutzer hat Zugriff auf alle AWS Dienste und Ressourcen im Konto. Als bewährte Methode zur Gewährleistung der Sicherheit sollten Sie den [administrativen Zugriff einem administrativen Benutzer zuweisen](#) und nur den Root-Benutzer verwenden, um [Aufgaben auszuführen, die Root-Benutzerzugriff erfordern](#).

## Erstellen eines Data Warehouse mit Amazon Redshift Serverless

Wenn Sie sich zum ersten Mal bei der Amazon-Redshift-Serverless-Konsole anmelden, werden Sie aufgefordert, auf die Informationen zu den ersten Schritten zuzugreifen, die Sie zum Erstellen und Verwalten von Serverless-Ressourcen verwenden können. In diesem Handbuch werden Sie Serverless-Ressourcen unter Verwendung der Standardeinstellungen von Amazon Redshift Serverless erstellen.

Wenn Sie Ihre Einrichtung genauer kontrollieren möchten, wählen Sie **Customize settings** (Einstellungen anpassen) aus.

So nehmen Sie die Konfiguration mit Standardeinstellungen vor:

1. Melden Sie sich bei der Amazon Redshift Redshift-Konsole an AWS Management Console und öffnen Sie sie unter <https://console.aws.amazon.com/redshiftv2/>.

Wählen Sie **Testen Sie Amazon Redshift Serverless** aus.

2. Wählen Sie unter **Configuration (Konfiguration)** die Option **Use default settings** (Standardeinstellungen verwenden) aus. Amazon Redshift Serverless erstellt einen Standard-Namespace mit einer zugeordneten Standardarbeitsgruppe. Wählen Sie **Save configuration** (Konfiguration speichern) aus.

Der folgende Screenshot zeigt die Standardeinstellungen für Amazon Redshift Serverless.

# Get started with Amazon Redshift Serverless

To start using Amazon Redshift Serverless, set up your serverless data warehouse and create a database. You will receive \$0 credit towards your Redshift Serverless usage in this account.

## Configuration [Info](#)

### Use default settings

Default settings have been defined to help you get started. You can change them at any time later.

### Customize settings

Customize your settings for your specific needs.

## ▼ How it works



### Using the default settings

Amazon Redshift Serverless creates a default namespace and workgroup. This configuration uses the default settings and becomes active when you associate the default workgroup to the default namespace.



### Customizing the settings

Amazon Redshift Serverless creates a default namespace and workgroup. This configuration becomes active when you associate the default workgroup to the default namespace.

## Namespace [Info](#)

Namespace is a collection of database objects and users. Data properties include database name and password, permissions, and encryption and security.

**⚠** Your data is encrypted by default with an AWS owned key. To choose a different key, choose **Customize settings**.

Namespace name  
default

### Database name and password

Database name  
dev

Admin user credentials  
Created based on IAM credentials

### Permissions

Default IAM role

<arn:aws:iam::694237154968:role/service-role/AmazonRedshift-CommandsAccessRole-20220809T211900>

3. Nachdem die Einrichtung abgeschlossen ist, wählen Sie Continue (Weiter), um zu Serverless Dashboard zu wechseln. Wie Sie sehen, sind die Serverless-Arbeitsgruppe und der Serverless-Namespace verfügbar.

**Serverless dashboard** [Info](#)

**Namespace overview** [Info](#)  
Namespace data from your account

Total snapshots	Datashares in my account	Datashares requiring authorization	Datashares fr
0	0	0	0

**Namespaces / Workgroups** [Info](#)

Namespace	Status	Workgroup	Status
default	✓ Available	default	✓ Available

## Laden von Beispieldaten

Nachdem Sie Ihr Data Warehouse mit Amazon Redshift Serverless eingerichtet haben, können Sie den Amazon Redshift Query Editor v2 verwenden, um Beispieldaten zu laden.

1. Um Query Editor v2 über die Amazon-Redshift-Serverless-Konsole zu starten, wählen Sie Daten abfragen aus. Wenn Sie den Abfrage-Editor v2 über die Amazon-Redshift-Serverless-Konsole aufrufen, wird er in einer neuen Browser-Registerkarte geöffnet. Der Abfrage-Editor v2 stellt eine Verbindung von Ihrem Clientcomputer mit der Amazon-Redshift-Serverless-Umgebung her.

Amazon Redshift Serverless

**Serverless dashboard** [Info](#)

[↻](#) **Query data** [↗](#) [Create workgroup](#)





2. Wenn Sie den Abfrage-Editor v2 zum ersten Mal starten, müssen Sie die AWS KMS Verschlüsselung konfigurieren, bevor Sie fortfahren können. Optional können Sie auch das URI für einen S3-Bucket angeben, um später Daten zu laden. Wählen Sie anschließend Konto konfigurieren aus.

## Configure account

To start using Redshift query editor v2, configure the account.


### AWS KMS encryption

Your data is encrypted by default with a key that AWS owns and manages for you. To choose a different key, customize your encryption settings.

 You can't change the AWS KMS encryption settings after configuring your account. 

Customize encryption settings (advanced)

### S3 bucket - optional

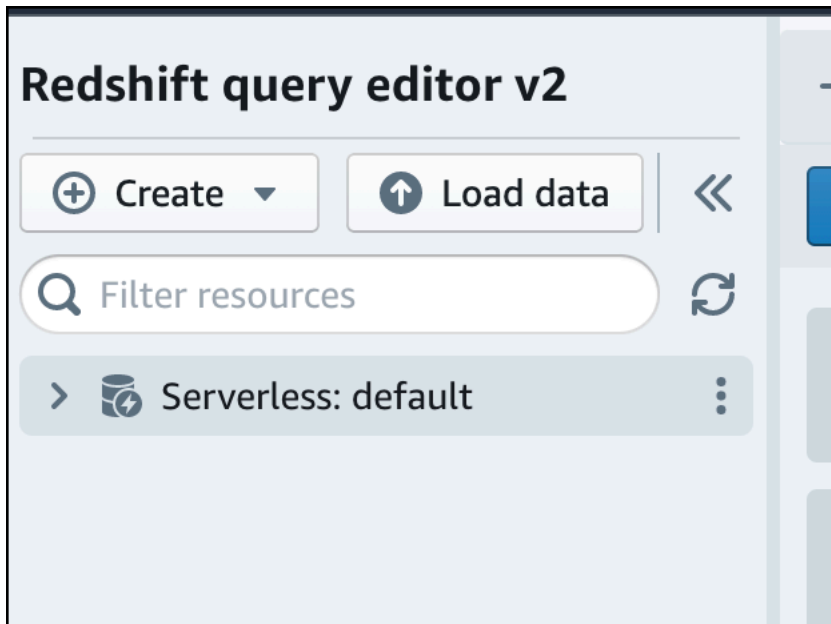
Enter an Amazon S3 path for this account. This path can be used for features such as load data from a local file store. Account users and the S3 bucket must be configured with the proper permissions to secure user's data. [Learn more](#) 

S3 URI

[Cancel](#) [Configure account](#)

Informationen zur Konfiguration von Query Editor v2, einschließlich der erforderlichen Berechtigungen, finden Sie unter [Konfigurieren Ihres AWS-Konto](#) im Amazon-Redshift-Verwaltungshandbuch.

3. Um eine Verbindung zu einer Arbeitsgruppe herzustellen, wählen Sie den Namen der Arbeitsgruppe im Strukturansichtsbereich aus.



4. Wenn Sie in Query Editor v2 zum ersten Mal eine Verbindung zu einer neuen Arbeitsgruppe herstellen, müssen Sie den Authentifizierungstyp auswählen, der für die Verbindung zur Arbeitsgruppe verwendet werden soll. Lassen Sie für diese Anleitung die Option Verbundbenutzer ausgewählt und wählen Sie Verbindung erstellen aus.

## Connect to default ✕

Authentication [Learn more](#)

**Federated user**  
The query editor v2 generates a temporary password to connect to the database.

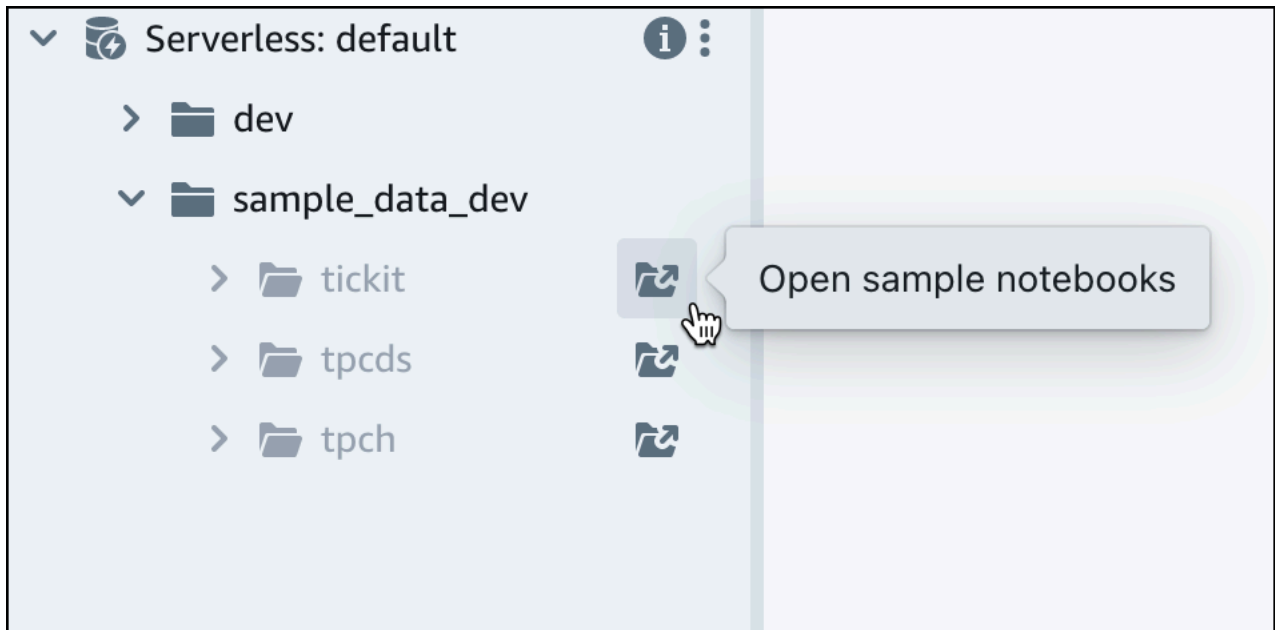
**Database user name and password**  
Provide a database user and password for the database that you are connecting to. The query editor v2 stores your credentials in AWS Secrets Manager on your behalf.

Database

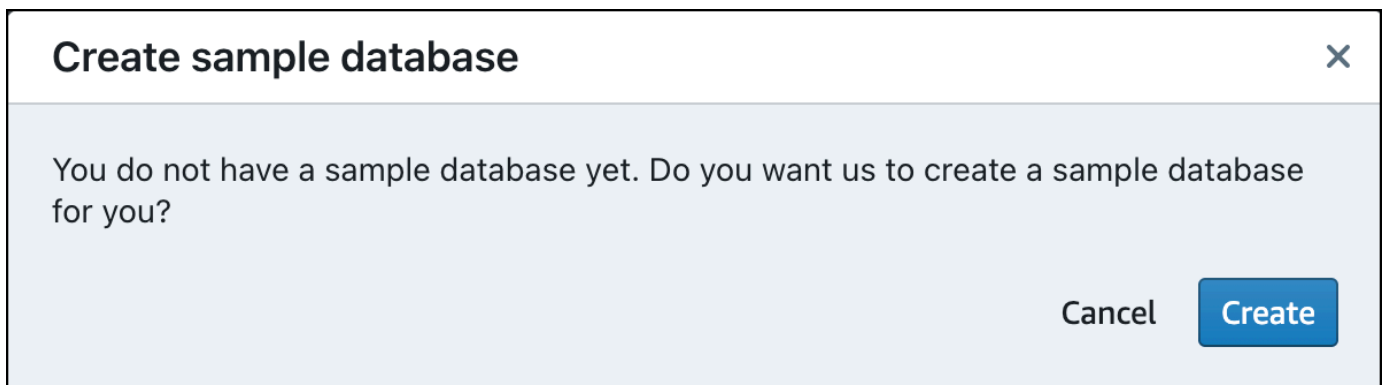
The database name must be 1-64 characters. Valid characters are lowercase alphanumeric characters.

Sobald Sie verbunden sind, können Sie Beispieldaten aus Amazon Redshift Serverless oder aus einem Amazon-S3-Bucket laden.

- Erweitern Sie unter der Standardarbeitsgruppe von Amazon Redshift Serverless die Datenbank `sample_data_dev`. Es gibt drei Beispielschemata, die drei Beispieldatensätzen entsprechen, die Sie in die Amazon-Redshift-Serverless-Datenbank laden können. Wählen Sie den Beispieldatensatz, den Sie laden möchten, und dann Beispiel-Notebooks öffnen aus.



6. Wenn Sie zum ersten Mal Daten laden, fordert Query Editor v2 Sie auf, eine Beispieldatenbank zu erstellen. Wählen Sie Erstellen.



## Ausführen von Beispielabfragen

Nachdem Sie Amazon Redshift Serverless eingerichtet haben, können Sie einen Beispieldatensatz in Amazon Redshift Serverless verwenden. Amazon Redshift Serverless lädt den Beispieldatensatz, z. B. den Tickit-Datensatz, automatisch und Sie können die Daten sofort abfragen.

- Sobald Amazon Redshift Serverless mit dem Laden der Beispieldaten fertig ist, werden alle Beispielabfragen in den Editor geladen. Sie können Alle auswählen, um alle Abfragen aus den Beispiel-Notebooks auszuführen.

The screenshot shows the Amazon Redshift console interface. At the top, the query title is "Sales per event". Below it, there are controls for "Run" and "Limit 100". The SQL query is displayed in a text area:

```

1 SET search_path to tickit;
2 SELECT eventname, total_price
3 FROM (SELECT eventid, total_price, ntile(1000) over(order by total_price desc) as percentile
4       FROM (SELECT eventid, sum(pricepaid) total_price
5             FROM tickit.sales
6             GROUP BY eventid)) Q, tickit.event E
7 WHERE Q.eventid = E.eventid
8      AND percentile = 1
9 ORDER BY total_price desc;

```

Below the query, there are tabs for "Result 1" and "Result 2 (9)". The "Result 2 (9)" tab is active, showing a table with 9 rows. The table has two columns: "eventname" and "total\_price". The data is as follows:

eventname	total_price
Adriana Lecouvreur	51846
Janet Jackson	51049
Phantom of the Opera	50301
The Little Mermaid	49956
Citizen Cope	49823
Sevendust	48020
Electra	47883
Mary Poppins	46780
Live	46661

At the bottom right of the results area, it says "Elapsed time: 401 ms Total rows: 9".

Sie können die Ergebnisse auch als JSON- oder CSV-Datei exportieren oder die Ergebnisse in einem Diagramm anzeigen.

This screenshot shows the same Amazon Redshift console interface as the previous one, but with additional features visible. The "Export" button (with a download icon) and the "Chart" toggle (which is currently turned on) are circled in red. Below the query, there is a "Structure" panel on the left with a "+ Trace" button. A "Traces" panel is open, showing "trace 0" with a "Type" of "Line". The "X" axis is set to "Data inlined in figure" and the "Y" axis is set to "Choose data...". To the right of the traces, there is a plot area with a grid and a blue line graph. The plot area has a title "Click to enter Plot title" and a y-axis label "Click to enter Y axis title". The y-axis has tick marks at 2, 4, 6, and 8. The plot area is currently empty, showing only the grid and the blue line.

Sie können Daten auch aus einem Amazon-S3-Bucket laden. Weitere Informationen hierzu finden Sie unter [the section called “Laden von Daten aus Amazon S3”](#).

# Laden von Daten aus Amazon S3

Nachdem Sie Ihr Data Warehouse erstellt haben, können Sie Daten aus Amazon S3 laden.

An diesem Punkt verfügen Sie über eine Datenbank namens dev. Als Nächstes legen Sie Tabellen in der Datenbank an, laden Daten in die Tabellen hoch und führen testweise eine Abfrage durch. Die Beispieldaten werden der Einfachheit halber in einem Amazon-S3-Bucket bereitgestellt.

1. Vor dem Laden von Daten aus Amazon S3 müssen Sie zunächst eine IAM-Rolle mit den erforderlichen Berechtigungen erstellen und Ihrem Serverless-Namespace hinzufügen. Wählen Sie dazu im Navigationsmenü die Option Namespace-Konfiguration und dann Sicherheit und Verschlüsselung. Wählen Sie IAM-Rollen verwalten.

The screenshot shows the Amazon Redshift console interface for a namespace named 'default'. The 'Security and encryption' tab is selected and highlighted with a red circle. Below the navigation tabs, the 'Workgroup name' section is visible, showing the workgroup name as 'default' and its status as 'Available'.

General information	
Namespace default	Status ✔ Available
Namespace ID example-namespace-id	Date created March 02, 2023, 12:11 (UTC-08:00)
Namespace ARN 📄 example-namespace-arn	Storage used 18.9 GB

**Workgroup name**  
Set up compute resources for your workgroup.

Workgroup default	Status ✔ Available
----------------------	-----------------------

2. Erweitern Sie das Menü IAM-Rollen verwalten und wählen Sie IAM-Rolle erstellen aus.

# Manage IAM roles

## Permissions

**i** Associate an IAM role so that your serverless endpoint can LOAD and UNLOAD data. You can create an IAM role as the default for this configuration that has the [AmazonRedshiftAllCommandsFullAccess](#) policy attached. This policy includes permissions to run SQL commands to COPY, UNLOAD, and query data with Amazon Redshift Serverless. This policy also grants permissions to run SELECT statements for related services, such as Amazon S3, Amazon CloudWatch logs, Amazon SageMaker, and AWS Glue. You won't be able to run these SQL commands without an IAM role attached to your namespace.

## Associated IAM roles (1)

Create, associate, or remove an IAM role. You can associate up to 50 IAM roles. You can also choose an IAM role and set it as the default.

Set default ▼

Manage IAM roles ▲



Search for associa

Associate IAM roles

Create IAM role

Remove IAM roles

or role type

< 1 >



IAM roles [↗](#)



Status



Role  
type



3. Wählen Sie die Ebene des S3-Bucket-Zugriffs aus, die Sie dieser Rolle gewähren möchten, und wählen Sie IAM-Rolle als Standard erstellen aus.

## Create the default IAM role ✕

**i** Associate an IAM role so that your serverless endpoint can LOAD and UNLOAD data. You can create an IAM role as the default for this configuration that has the [AmazonRedshiftAllCommandsFullAccess](#) policy attached. This policy includes permissions to run SQL commands to COPY, UNLOAD, and query data with Amazon Redshift Serverless. This policy also grants permissions to run SELECT statements for related services, such as Amazon S3, Amazon CloudWatch logs, Amazon SageMaker, and AWS Glue. You won't be able to run these SQL commands without an IAM role attached to your namespace.

Specify an S3 bucket for the IAM role to access  
To create a new bucket, [visit S3](#)

No additional S3 bucket  
Create the IAM role without specifying S3 buckets.

Any S3 bucket  
Allow users that have access to your Redshift Serverless data to also access any S3 bucket and its contents in your AWS account.

Specific S3 buckets  
Specify one or more S3 buckets that the IAM role being created has permission to access.

4. Wählen Sie Änderungen speichern aus. Sie können jetzt Beispieldaten aus Amazon S3 laden.

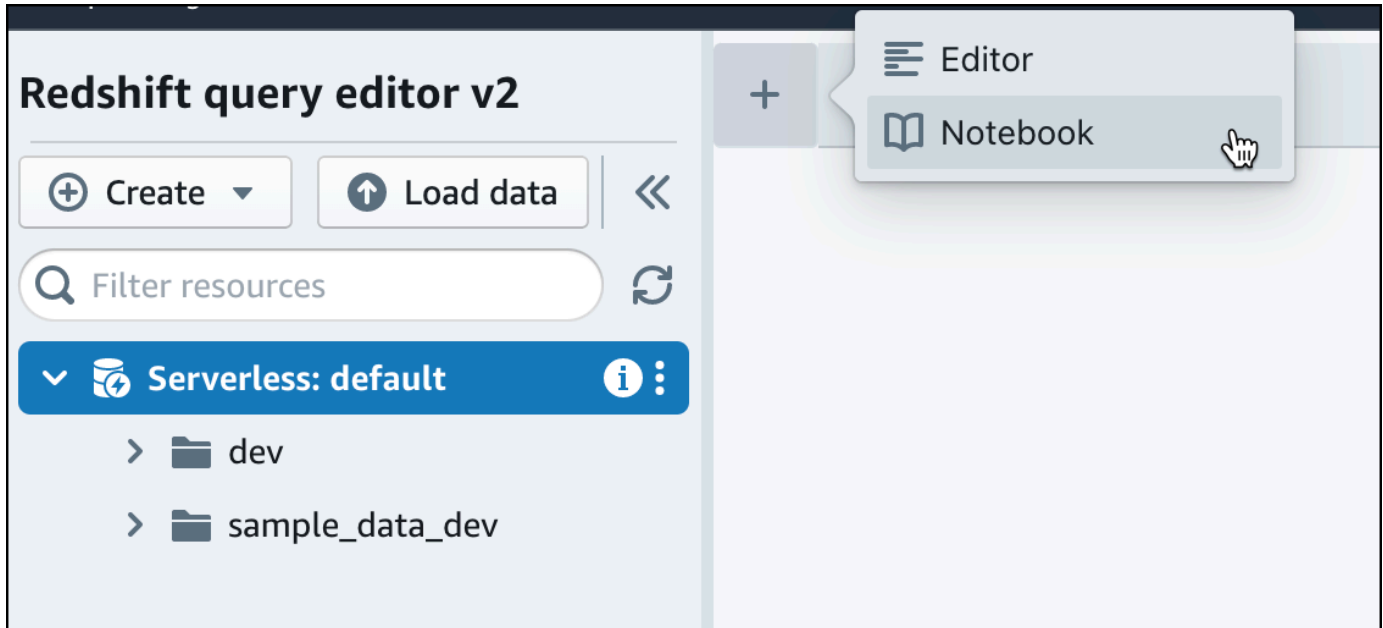
In den folgenden Schritten werden Daten in einem öffentlichen Amazon-Redshift-S3-Bucket verwendet, Sie können jedoch dieselben Schritte unter Verwendung Ihres eigenen S3-Buckets und eigener SQL-Befehle wiederholen.



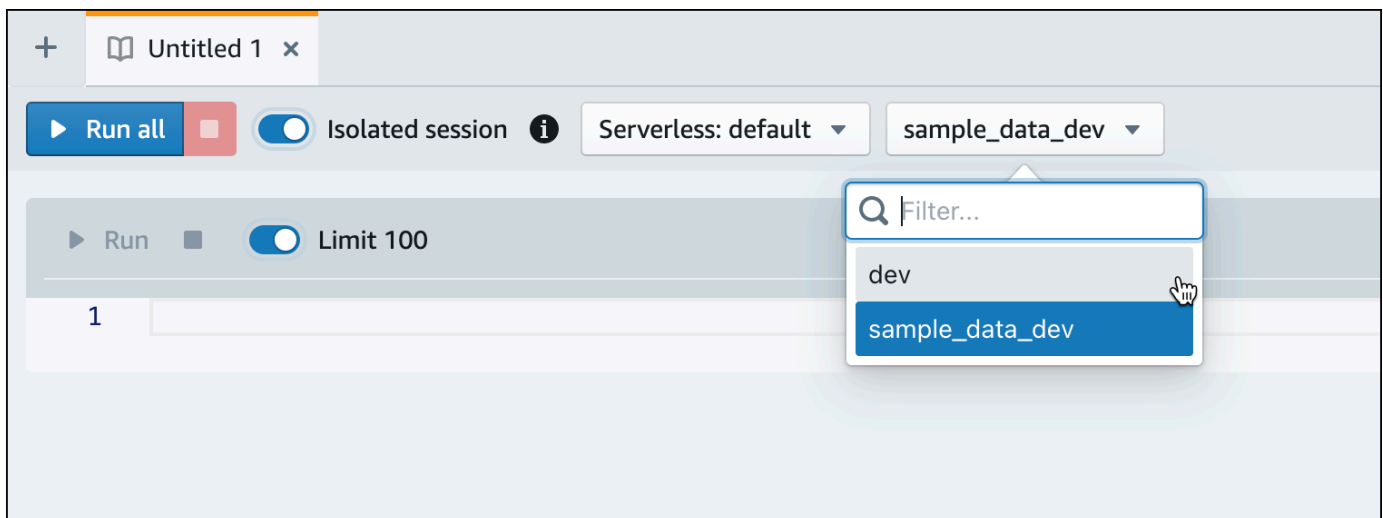
## Laden von Beispieldaten aus Amazon S3

1. Wählen Sie in Query Editor v2 „  
+

hinzufügen“ und dann Notebook aus, um ein neues SQL-Notebook zu erstellen.



2. Wechseln Sie zur dev-Datenbank.



3. Erstellen Sie Tabellen.

Wenn Sie Query Editor v2 verwenden, kopieren Sie die folgenden Create-Table-Anweisungen und führen Sie sie aus, um Tabellen in der dev-Datenbank zu erstellen. Weitere Informationen

zur Syntax finden Sie unter [CREATE TABLE](#) im Datenbankentwicklerhandbuch zu Amazon Redshift.

```
create table users(  
userid integer not null distkey sortkey,  
username char(8),  
firstname varchar(30),  
lastname varchar(30),  
city varchar(30),  
state char(2),  
email varchar(100),  
phone char(14),  
likesports boolean,  
liketheatre boolean,  
likeconcerts boolean,  
likejazz boolean,  
likeclassical boolean,  
likeopera boolean,  
likerock boolean,  
likevegas boolean,  
likebroadway boolean,  
likemusicals boolean);
```

```
create table event(  
eventid integer not null distkey,  
venueid smallint not null,  
catid smallint not null,  
dateid smallint not null sortkey,  
eventname varchar(200),  
starttime timestamp);
```

```
create table sales(  
salesid integer not null,  
listid integer not null distkey,  
sellerid integer not null,  
buyerid integer not null,  
eventid integer not null,  
dateid smallint not null sortkey,  
qtysold smallint not null,  
pricepaid decimal(8,2),  
commission decimal(8,2),  
saletime timestamp);
```

#### 4. Erstellen Sie in Query Editor v2 eine neue SQL-Zelle in Ihrem Notebook.



#### 5. Verwenden Sie nun den Befehl COPY in Query Editor v2, um große Datensätze aus Amazon S3 oder Amazon DynamoDB in Amazon Redshift zu laden. Weitere Informationen zur COPY-Syntax finden Sie unter [COPY](#) im Datenbankentwicklerhandbuch zu Amazon Redshift.

Sie können den Befehl COPY mit Beispieldaten ausführen, die in einem öffentlichen S3-Bucket verfügbar sind. Führen Sie die folgenden SQL-Befehle in Query Editor v2 aus.

```
COPY users
FROM 's3://redshift-downloads/ticket/allusers_pipe.txt'
DELIMITER '|'
TIMEFORMAT 'YYYY-MM-DD HH:MI:SS'
IGNOREHEADER 1
REGION 'us-east-1'
IAM_ROLE default;

COPY event
FROM 's3://redshift-downloads/ticket/allevnts_pipe.txt'
DELIMITER '|'
TIMEFORMAT 'YYYY-MM-DD HH:MI:SS'
IGNOREHEADER 1
REGION 'us-east-1'
IAM_ROLE default;

COPY sales
FROM 's3://redshift-downloads/ticket/sales_tab.txt'
DELIMITER '\t'
TIMEFORMAT 'MM/DD/YYYY HH:MI:SS'
IGNOREHEADER 1
REGION 'us-east-1'
IAM_ROLE default;
```

- Erstellen Sie nach dem Laden der Daten eine weitere SQL-Zelle in Ihrem Notebook und probieren Sie einige Beispielabfragen aus. Weitere Informationen zur Verwendung des SELECT-Befehls finden Sie unter [SELECT](#) im Amazon-Redshift-Entwicklerhandbuch. Verwenden Sie Query Editor v2, um die Struktur und die Schemata der Beispieldaten zu verstehen.

```
-- Find top 10 buyers by quantity.
SELECT firstname, lastname, total_quantity
FROM (SELECT buyerid, sum(qtysold) total_quantity
      FROM sales
      GROUP BY buyerid
      ORDER BY total_quantity desc limit 10) Q, users
WHERE Q.buyerid = userid
ORDER BY Q.total_quantity desc;

-- Find events in the 99.9 percentile in terms of all time gross sales.
SELECT eventname, total_price
FROM (SELECT eventid, total_price, ntile(1000) over(order by total_price desc) as
      percentile
      FROM (SELECT eventid, sum(pricepaid) total_price
            FROM sales
            GROUP BY eventid)) Q, event E
WHERE Q.eventid = E.eventid
AND percentile = 1
ORDER BY total_price desc;
```

Nachdem Sie nun Daten geladen und einige Beispielabfragen ausgeführt haben, können Sie andere Bereiche von Amazon Redshift Serverless erkunden. In der folgenden Übersicht erfahren Sie mehr über die Verwendungsmöglichkeiten von Amazon Redshift Serverless.

- Sie können Daten aus einem Amazon-S3-Bucket laden. Weitere Informationen finden Sie unter [Laden von Daten aus Amazon S3](#).
- Sie können Query Editor v2 verwenden, um Daten aus einer lokalen zeichengetrennten Datei mit weniger als 5 MB zu laden. Weitere Informationen finden Sie unter [Laden von Daten aus einer lokalen Datei](#).
- Sie können eine Verbindung zu Amazon Redshift Serverless mit SQL-Tools von Drittanbietern mit dem JDBC- und ODBC-Treiber herstellen. Weitere Informationen finden Sie unter [Verbinden mit Amazon Redshift Serverless](#).

- Sie können die Amazon-Redshift-Daten-API auch verwenden, um eine Verbindung mit Amazon Redshift Serverless herzustellen. Weitere Informationen finden Sie unter [Verwenden der Amazon-Redshift-Daten-API](#).
- Sie können Ihre Daten in Amazon Redshift Serverless mit Redshift ML verwenden, um Machine-Learning-Modelle mit dem Befehl CREATE MODEL zu erstellen. Im [Tutorial: Erstellen von Kundenabwanderungsmodellen](#) erfahren Sie, wie Sie ein Redshift-ML-Modell erstellen.
- Sie können Daten aus einem Amazon S3 Data Lake abfragen, ohne Daten in Amazon Redshift Serverless laden zu müssen. Weitere Informationen finden Sie unter [Abfragen eines Data Lake](#).

# Von Amazon Redshift bereitgestellte Cluster

Wenn Sie Amazon Redshift zum ersten Mal verwenden, hilft Ihnen dieses Tutorial bei den ersten Schritten mit Amazon Redshift Redshift-Clustern. Informationen zu den ersten Schritten mit Redshift Serverless finden Sie unter. [Amazon Redshift Serverless](#)

Nachdem Sie die bereitgestellte Amazon Redshift Redshift-Konsole erstellt und sich dort angemeldet haben, können Sie Amazon Redshift Redshift-Objekte, einschließlich Cluster, Knoten und Datenbanken, erstellen und verwalten. Sie können mit einem SQL-Client auch Abfragen ausführen, Abfragen anzeigen und andere Operationen in der SQL Data Definition Language (DDL) und Data Manipulation Language (DML) ausführen.

## Important

Der Cluster, den Sie für diese Übung bereitstellen, wird in einer Live-Umgebung ausgeführt. Solange er läuft, fallen Gebühren für Sie an. AWS-Konto Informationen zu Preisen finden Sie auf der [Amazon-Redshift-Preisseite](#).

Um unnötige Kosten zu vermeiden, sollten Sie den Cluster löschen, wenn Sie damit fertig sind. Im letzten Schritt des Tutorials wird erklärt, wie das geht.

Melden Sie sich bei der Amazon Redshift Redshift-Konsole an AWS Management Console und öffnen Sie sie unter <https://console.aws.amazon.com/redshiftv2/>.

Wir empfehlen Ihnen, zunächst das Dashboard für bereitgestellte Cluster aufzurufen, um mit der Nutzung der Amazon Redshift Redshift-Konsole zu beginnen.

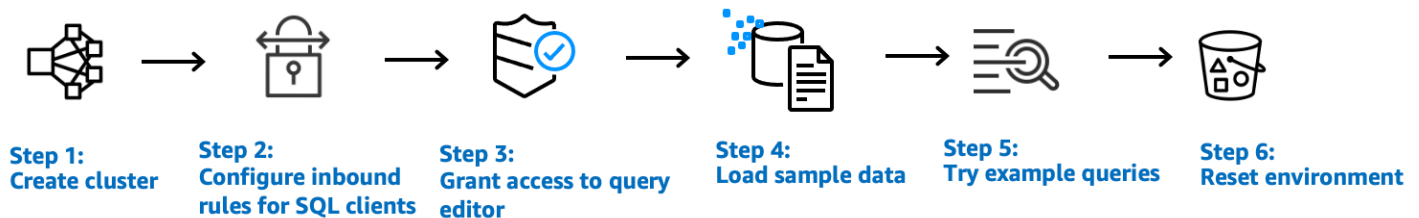
Abhängig von Ihrer Konfiguration werden die folgenden Elemente im Navigationsbereich der von Amazon Redshift bereitgestellten Konsole angezeigt:

- Redshift Serverless — Greifen Sie auf Daten zu und analysieren Sie sie, ohne die von Amazon Redshift bereitgestellten Cluster einrichten, optimieren und verwalten zu müssen.
- Dashboard für bereitgestellte Cluster — Sehen Sie sich die Liste der Cluster in Ihrem an AWS-Region, überprüfen Sie Cluster-Metriken und Abfrageübersicht, um Einblicke in Metrikdaten (wie CPU-Auslastung) und Abfrageinformationen zu erhalten. Damit können Sie feststellen, ob Ihre Leistungsdaten über einen bestimmten Zeitraum anomal sind.

- Cluster — Sehen Sie sich hier Ihre Clusterliste an AWS-Region, wählen Sie einen Cluster aus, um mit der Abfrage zu beginnen, oder führen Sie clusterbezogene Aktionen durch. Sie können von dieser Seite aus auch einen neuen Cluster erstellen.
- Abfrage-Editor — Führen Sie Abfragen für Datenbanken aus, die auf Ihrem Amazon Redshift Redshift-Cluster gehostet werden. Wir empfehlen, stattdessen den Query-Editor v2 zu verwenden.
- Query Editor v2 — Amazon Redshift Query Editor v2 ist eine separate webbasierte SQL-Client-Anwendung zum Verfassen und Ausführen von Abfragen in Ihrem Amazon Redshift Data Warehouse. Sie können Ihre Ergebnisse in Diagrammen darstellen und Abfragen mit anderen im Team teilen.
- Queries and loads (Abfragen und Ladevorgänge) – Rufen Sie Informationen als Referenz oder zur Fehlerbehebung auf, z. B. eine Liste der letzten Abfragen und den SQL-Text für jede Abfrage.
- Datashares – Als Administratoren eines Produzentenkontos können Sie entweder Konsumentenkonten zum Zugreifen auf Datashares autorisieren oder keinen Zugriff autorisieren. Um ein autorisiertes Datashare zu verwenden, kann ein Administrator eines Benutzerkontos das Datashare entweder einem ganzen AWS-Konto oder bestimmten Cluster-Namespaces in einem Konto zuordnen. Ein Administrator kann ein Datashare auch ablehnen.
- Zero-ETL-Integrationen — Verwalten Sie Integrationen, die Transaktionsdaten in Amazon Redshift verfügbar machen, nachdem sie in unterstützte Quellen geschrieben wurden.
- IAM Identity Center-Verbindungen — Konfigurieren Sie die Verbindung zwischen Amazon Redshift und IAM Identity Center.
- Configurations (Konfigurationen) – Stellen Sie über Java-Database-Connectivity(JDBC)- und Open-Database-Connectivity(ODBC)-Verbindungen eine Verbindung von SQL-Client-Tools zu Amazon-Redshift-Clustern her. Sie können auch einen von Amazon Redshift verwalteten Virtual Private Cloud (VPC)-Endpunkt einrichten. Dadurch wird eine private Verbindung hergestellt zwischen einer VPC, die auf dem Amazon-VPC-Service basiert, der einen Cluster enthält, und einer anderen VPC, in der ein Client-Tool ausgeführt wird.
- AWS Partnerintegration — Erstellen Sie eine Integration mit einem unterstützten Partner. AWS
- Advisor – Erhalten Sie spezifische Empfehlungen zu Änderungen, die Sie an Ihrem Amazon-Redshift-Cluster vornehmen können, um Ihre Optimierungen zu priorisieren.
- AWS Marketplace — Informieren Sie sich über andere Tools oder AWS Services, die mit Amazon Redshift funktionieren.
- Alarms (Alarmer) – Erstellen Sie Alarmer für Cluster-Metriken, um Leistungsdaten anzuzeigen und Metriken über einen von Ihnen festgelegten Zeitraum nachzuverfolgen.

- Events (Ereignisse) – Verfolgen Sie Ereignisse nach und rufen Sie Berichte mit Informationen wie dem Datum des Ereignisses, einer Beschreibung oder der Ereignisquelle ab.
- What's new (Neuerungen) – Zeigen Sie neue Funktionen und Produktaktualisierungen von Amazon Redshift an.

In diesem Tutorial führen Sie die folgenden Schritte durch:



## Themen

- [Melden Sie sich an für AWS](#)
- [Festlegen von Firewall-Regeln](#)
- [Schritt 1: Erstellen eines Amazon-Redshift-Beispielclusters](#)
- [Schritt 2: Regeln für eingehenden Datenverkehr für SQL-Clients konfigurieren](#)
- [Schritt 3: Gewähren Sie Zugriff auf einen SQL-Client und führen Sie Abfragen aus](#)
- [Schritt 4: Daten aus Amazon S3 in Amazon Redshift laden](#)
- [Schritt 5: Beispielabfragen mit dem Abfrage-Editor testen](#)
- [Schritt 6: Umgebung zurücksetzen](#)

## Melden Sie sich an für AWS

Wenn Sie noch keine haben AWS-Konto, melden Sie sich für eine an. Wenn Sie bereits ein Konto besitzen, können Sie diesen Schritt überspringen und Ihr vorhandenes Konto verwenden.

1. Öffnen Sie <https://portal.aws.amazon.com/billing/signup>.
2. Folgen Sie den Online-Anweisungen.

Bei der Anmeldung müssen Sie auch einen Telefonanruf entgegennehmen und einen Verifizierungscode über die Tasten eingeben.

Wenn Sie sich für eine anmelden AWS-Konto, Root-Benutzer des AWS-Kontos wird eine erstellt. Der Root-Benutzer hat Zugriff auf alle AWS-Services und Ressourcen des Kontos. Aus



Sicherheitsgründen sollten Sie einem Benutzer Administratorzugriff zuweisen und nur den Root-Benutzer verwenden, um [Aufgaben auszuführen, für die Root-Benutzerzugriff erforderlich](#) ist.

## Festlegen von Firewall-Regeln

### Note

In diesem Tutorial wird davon ausgegangen, dass Ihr Cluster den Standardport 5439 verwendet und der Amazon Redshift Query Editor v2 zur Ausführung von SQL-Befehlen verwendet werden kann. Es geht nicht auf Details zu Netzwerkkonfigurationen oder zur Einrichtung eines SQL-Clients ein, die in Ihrer Umgebung erforderlich sein könnten.

In einigen Umgebungen geben Sie einen Port an, wenn Sie Ihren Amazon Redshift Redshift-Cluster starten. Des Weiteren erstellen Sie eine Zugangsregel für eingehenden Datenverkehr in einer Sicherheitsgruppe, die den Zugriff auf den Cluster über den Port regelt.

Wenn der Client-Computer durch eine Firewall geschützt ist, müssen Sie wissen, welcher Port offen ist. Über diesen offenen Port können Sie von einem SQL-Client-Tool eine Verbindung zum Cluster herstellen und Abfragen ausführen. Sollten Sie keinen offenen Port kennen, bitten Sie jemanden, der sich mit den Firewall-Regeln in Ihrem Netzwerk auskennt, einen offenen Port in der Firewall zu ermitteln.

Standardmäßig verwendet Amazon Redshift Port 5439. Eine Verbindung ist aber nur möglich, wenn dieser Port in der Firewall auch geöffnet ist. Sie können die Portnummer für den Amazon-Redshift-Cluster nach der Erstellung nicht mehr ändern. Stellen Sie daher sicher, dass Sie einen offenen Port angeben, der in Ihrer Umgebung beim Startvorgang funktioniert.

## Schritt 1: Erstellen eines Amazon-Redshift-Beispielclusters


In diesem Tutorial gehen Sie durch den Prozess zur Erstellung eines Amazon Redshift Redshift-Clusters mit einer Datenbank. Dann laden Sie einen Datensatz aus Amazon S3 in Tabellen Ihrer Datenbank. Anhand dieses Beispiel-Clusters können Sie den Dienst Amazon Redshift auswerten.

Bevor Sie mit der Einrichtung eines Amazon Redshift Redshift-Clusters beginnen, stellen Sie sicher, dass Sie alle erforderlichen Voraussetzungen wie [Melden Sie sich an für AWS](#) und [Festlegen von Firewall-Regeln](#) erfüllen.

Für jeden Vorgang, der auf Daten von einer anderen AWS Ressource zugreift, benötigt Ihr Cluster die Erlaubnis, in Ihrem Namen auf die Ressource und die Daten auf der Ressource zuzugreifen. Ein Beispiel ist die Verwendung eines SQL COPY-Befehls zum Laden von Daten aus Amazon Simple Storage Service (Amazon S3). Sie stellen diese Berechtigungen mithilfe von AWS Identity and Access Management (IAM) bereit. Sie können dies über eine IAM-Rolle tun, die Sie erstellen und Ihrem Cluster zuordnen. Weitere Informationen zu Anmeldeinformationen und Zugriffsberechtigungen finden Sie unter [Anmeldeinformationen und Zugriffsberechtigungen](#) im Amazon Redshift Database Developer Guide.


So erstellen Sie einen Amazon-Redshift-Cluster

1. Melden Sie sich bei der Amazon Redshift Redshift-Konsole an AWS Management Console und öffnen Sie sie unter <https://console.aws.amazon.com/redshiftv2/>.

 **Important**

Wenn Sie IAM-Benutzeranmeldeinformationen verwenden, stellen Sie sicher, dass Sie über die erforderlichen Berechtigungen verfügen, um die Cluster-Operationen durchzuführen. Weitere Informationen finden Sie unter [Sicherheit in Amazon Redshift](#) im Amazon Redshift Management Guide.

2. Wählen Sie auf der AWS Konsole den AWS-Region Ort aus, an dem Sie den Cluster erstellen möchten.
3. Wählen Sie im Navigationsmenü Clusters (Cluster) und dann Create cluster (Cluster erstellen) aus. Die Seite Create Cluster (Cluster erstellen) wird angezeigt.
4. Geben Sie im Bereich Cluster configuration (Cluster-Konfiguration) Werte für Cluster identifier (Cluster-ID), Node type (Knotentyp) und Nodes (Knoten) an:
  - Cluster Identifier (Cluster-ID): Geben Sie für dieses Tutorial **examplecluster** ein. Diese ID muss eindeutig sein. Die ID muss aus 1—63 Zeichen bestehen und darf nur die Buchstaben a–z (nur Kleinschreibung) und - (Bindestriche) enthalten.
  - Wählen Sie eine der folgenden Methoden zur Bestimmung der Größe Ihres Clusters aus:

 **Note**

Im folgenden Schritt wird davon ausgegangen AWS-Region , dass ein System RA3-Knotentypen unterstützt. Eine Liste von AWS-Regionen , die RA3-Knotentypen unterstützen, finden Sie unter [Übersicht über RA3-Knotentypen](#) im Amazon-Redshift-

Verwaltungshandbuch. Weitere Informationen über die Knotenspezifikationen für die einzelnen Knotentypen und -größen finden Sie unter [Details zu Knotentypen](#).

- Wenn Sie nicht wissen, wie groß Ihr Cluster sein sollte, wählen Sie Help me choose (Hilfe bei der Auswahl) aus. Dadurch wird ein Größenrechner geöffnet, der Ihnen Fragen zur Größe und zu den Abfrageeigenschaften der Daten stellt, die Sie in Ihrem Data Warehouse speichern möchten.

Wenn Sie die erforderliche Größe Ihres Clusters kennen (d. h. den Knotentyp und die Anzahl der Knoten), wählen Sie I'll choose (Ich entscheide) aus. Wählen Sie den Node type (Knotentyp) und die Anzahl der Nodes (Knoten) aus, um die Größe Ihres Clusters für den Machbarkeitsnachweis zu bestimmen.

Wählen Sie für dieses Tutorial ra3.4xlarge als Knotentyp und 2 für Anzahl der Knoten.

Wenn eine AZ-Konfiguration verfügbar ist, wählen Sie Single-AZ.

- Wählen Sie unter Sample data (Beispieldaten) Load sample data (Beispieldaten laden) aus, um den Beispieldatensatz zu verwenden, den Amazon Redshift bereitstellt. Amazon Redshift lädt den Beispieldatensatz Tickit in die standardmäßige dev-Datenbank und das public-Schema.
5. Geben Sie im Bereich Datenbankkonfiguration einen Wert für Administrator-Benutzername ein. Wählen Sie für Administratorpasswort eine der folgenden Optionen aus:
- Ein Passwort erstellen – Verwendung eines von Amazon Redshift generierten Passworts.
  - Administratorpasswort manuell hinzufügen – Verwendung Ihres eigenen Passworts.
  - Administratoranmeldedaten verwalten in AWS Secrets Manager — Amazon Redshift verwendet AWS Secrets Manager, um Ihr Administratorkennwort zu generieren und zu verwalten. Für AWS Secrets Manager die Generierung und Verwaltung Ihres Passworts fällt eine Gebühr an. Informationen zu den Preisen für AWS Secrets Manager finden Sie unter [AWS Secrets Manager – Preise](#).

Verwenden Sie für dieses Tutorial folgende Werte:

- Admin user name (Administratorbenutzername): Geben Sie **awsuser** ein.
- Admin user password (Administratorpasswort): Geben Sie einen Wert für das Passwort ein.

6. Erstellen Sie für dieses Tutorial eine IAM-Rolle und legen Sie sie als Standard für Ihren Cluster fest, wie nachfolgend beschrieben. Für einen Cluster kann nur eine Standard-IAM-Rolle festgelegt werden.
  - a. Wählen Sie unter Cluster permissions (Cluster-Berechtigungen) bei Manage IAM roles (IAM-Rollen verwalten) die Option Create IAM role (IAM-Rolle erstellen) aus.
  - b. Geben Sie einen Amazon S3 Bucket an, auf den die IAM-Rolle zugreifen soll, indem Sie eine der folgenden Methoden verwenden:
    - Wählen Sie No additional Amazon S3 bucket (Kein zusätzlicher Amazon S3 Bucket) aus, damit die erstellte IAM-Rolle nur auf die Amazon S3 Buckets zugreifen kann, die als `redshift` benannt sind.
    - Wählen Sie Any Amazon S3 bucket (Beliebiger Amazon S3 Bucket) aus, damit die erstellte IAM-Rolle auf alle Amazon S3 Buckets zugreifen kann.
    - Wählen Sie Specific Amazon S3 buckets (Bestimmte Amazon S3 Buckets) aus, um einen oder mehrere Amazon S3 Buckets anzugeben, auf die die erstellte IAM-Rolle Zugriff hat. Wählen Sie dann einen oder mehrere Amazon S3 Buckets aus der Tabelle aus.
  - c. Wählen Sie Create IAM role as default (IAM-Rolle als Standard erstellen) aus. Amazon Redshift erstellt automatisch die Rolle und legt sie als Standard für Ihren Cluster fest.

Da Sie Ihre IAM-Rolle von der Konsole aus erstellt haben, ist ihr die Richtlinie `AmazonRedshiftAllCommandsFullAccess` angefügt. Dadurch kann Amazon Redshift Daten von Amazon-Ressourcen in Ihrem IAM-Konto kopieren, laden, abfragen und analysieren.

Informationen zur Verwaltung der Standard-IAM-Rolle für einen Cluster finden Sie unter [Creating an IAM role as default for Amazon Redshift im Amazon Redshift Management Guide](#).

7. (Optional) Deaktivieren Sie im Bereich Additional configurations (Zusätzliche Konfigurationen) die Option Use defaults (Standardwerte verwenden), um die Einstellungen Network and security (Netzwerk und Sicherheit), Database configuration (Datenbankkonfiguration), Maintenance (Wartung), Monitoring (Überwachung) und Backup anzupassen.

In manchen Fällen können Sie Ihren Cluster mit der Option Load sample data (Beispieldaten laden) erstellen. Dabei empfiehlt es sich möglicherweise, erweitertes Amazon-VPC-Routing zu aktivieren. In diesem Fall benötigt der Cluster in Ihrer Virtual Private Cloud (VPC) Zugriff auf den Amazon-S3-Endpunkt, damit Daten geladen werden können.

Um den Cluster öffentlich zugänglich zu machen, haben Sie zwei Möglichkeiten. Sie können eine NAT-Adresse (Network Address Translation) in Ihrer VPC konfigurieren, damit der Cluster auf das Internet zugreifen kann. Oder Sie können einen Amazon-S3-VPC-Endpoint in Ihrer VPC konfigurieren. Weitere Informationen zu erweitertem Amazon VPC-Routing finden Sie unter [Verbessertes Amazon VPC-Routing](#) im Amazon Redshift Management Guide.

8. Wählen Sie Cluster erstellen. Warten Sie, bis Ihr Cluster mit dem **Available** Status auf der Cluster-Seite erstellt wurde.

## Schritt 2: Regeln für eingehenden Datenverkehr für SQL-Clients konfigurieren

### Note

Wir empfehlen Ihnen, diesen Schritt zu überspringen und mit dem Amazon Redshift Query Editor v2 auf Ihren Cluster zuzugreifen.

Im weiteren Verlauf dieses Tutorials greifen Sie aus einer Virtual Private Cloud (VPC) auf Grundlage des Amazon-VPC-Service heraus auf Ihren Cluster zu. Wenn Sie einen SQL-Client von außerhalb Ihrer Firewall für den Zugriff auf den Cluster verwenden, müssen Sie jedoch den eingehenden Zugriff gewähren.

So überprüfen Sie Ihre Firewall und gewähren eingehenden Zugriff auf Ihren Cluster:

1. Überprüfen Sie Ihre Firewall-Regeln, wenn auf Ihren Cluster von außerhalb einer Firewall zugegriffen werden muss. Ihr Client kann beispielsweise eine Amazon Elastic Compute Cloud (Amazon EC2)-Instance oder ein externer Computer sein.

Weitere Informationen zu Firewallregeln finden Sie unter [Sicherheitsgruppenregeln](#) im Amazon EC2 EC2-Benutzerhandbuch.

2. Um von einem externen Amazon-EC2-Client aus darauf zuzugreifen, fügen Sie der an Ihren Cluster angefügten Sicherheitsgruppe, die eingehenden Datenverkehr zulässt, eine Eingangsregel hinzu. Sie fügen Amazon-EC2-Sicherheitsgruppenregeln in der Amazon-EC2-Konsole hinzu. So ermöglicht die CIDR/IP 192.0.2.0/24 Clients in diesem IP-Adressbereich eine Verbindung zu Ihrem Cluster. Finden Sie die richtige CIDR/IP für Ihre Umgebung heraus.

## Schritt 3: Gewähren Sie Zugriff auf einen SQL-Client und führen Sie Abfragen aus

Um Datenbanken abzufragen, die von Ihrem Amazon Redshift Redshift-Cluster gehostet werden, haben Sie mehrere Optionen für SQL-Clients. Dazu zählen:

- Connect zu Ihrem Cluster her und führen Sie Abfragen mit dem Amazon Redshift Query Editor v2 aus.

Wenn Sie den Abfrage-Editor v2 verwenden, müssen Sie keine SQL-Client-Anwendung herunterladen und einrichten. Sie starten den Amazon Redshift Query Editor v2 von der Amazon Redshift Redshift-Konsole aus.

- Stellen Sie mithilfe von RSQL eine Connect zu Ihrem Cluster her. Weitere Informationen finden Sie unter [Connecting with Amazon Redshift RSQL](#) im Amazon Redshift Management Guide.
- Connect Sie über ein SQL-Client-Tool wie SQL Workbench/J eine Verbindung zu Ihrem Cluster her. Weitere Informationen finden Sie unter [Connect zu Ihrem Cluster mithilfe von SQL Workbench/J](#) im Amazon Redshift Management Guide.

In diesem Tutorial wird Amazon Redshift Query Editor v2 als einfache Methode zum Ausführen von Abfragen in Datenbanken verwendet, die von Ihrem Amazon Redshift Redshift-Cluster gehostet werden. Nachdem Sie Ihren Cluster erstellt haben, können Sie sofort Abfragen ausführen. Einzelheiten zu Überlegungen bei der Verwendung des Amazon Redshift-Abfrage-Editors v2 finden Sie unter [Überlegungen bei der Arbeit mit dem Abfrage-Editor v2](#) im Amazon Redshift Management Guide.

### Gewähren des Zugriffs auf Abfrage-Editor v2

Wenn ein Administrator den Abfrage-Editor v2 zum ersten Mal für Sie konfiguriert, wählt er den aus AWS-Konto, der zum Verschlüsseln der AWS KMS key Query Editor v2-Ressourcen verwendet wird. Zu den Ressourcen des Amazon Redshift Query Editor v2 gehören gespeicherte Abfragen, Notizbücher und Diagramme. Standardmäßig werden die Ressourcen mit einem AWS -eigenen Schlüssel verschlüsselt. Alternativ kann ein Administrator einen vom Kunden verwalteten Schlüssel verwenden, indem er auf der Konfigurationsseite den Amazon-Ressourcennamen (ARN) für den Schlüssel auswählt. Nachdem Sie ein Konto konfiguriert haben, können die AWS KMS Verschlüsselungseinstellungen nicht mehr geändert werden. Weitere Informationen finden Sie unter [Konfiguration Ihres AWS-Konto](#) im Amazon Redshift Management Guide.

Um den Abfrage-Editor v2 aufzurufen, benötigen Sie eine Berechtigung. Ein Administrator kann eine der AWS verwalteten Richtlinien für Amazon Redshift Query Editor v2 an die IAM-Rolle oder den IAM-Benutzer anhängen, um Berechtigungen zu erteilen. Diese AWS verwalteten Richtlinien verfügen über verschiedene Optionen, mit denen gesteuert wird, wie das Markieren von Ressourcen die gemeinsame Nutzung von Abfragen ermöglicht. Sie können zum Zuweisen von IAM-Richtlinien die IAM-Konsole (<https://console.aws.amazon.com/iam/>) verwenden. Weitere Informationen zu diesen Richtlinien finden Sie unter [Accessing the Query Editor v2](#) im Amazon Redshift Management Guide.

Sie können auch Ihre eigene Richtlinie erstellen, basierend auf den zulässigen und verweigerten Berechtigungen in den bereitgestellten verwalteten Richtlinien. Wenn Sie den IAM-Konsolenrichtlinien-Editor verwenden, um Ihre eigene Richtlinie zu erstellen, wählen Sie SQL Workbench als Service aus, für den Sie die Richtlinie im visuellen Editor erstellen. Der Abfrage-Editor v2 verwendet den Servicenamen AWS SQL Workbench im Visual Editor und im IAM Policy Simulator.

Weitere Informationen finden Sie unter [Arbeiten mit dem Abfrage-Editor v2](#) im Amazon-Redshift-Verwaltungshandbuch.

## Schritt 4: Daten aus Amazon S3 in Amazon Redshift laden

Nachdem Sie Ihren Cluster erstellt haben, können Sie Daten aus Amazon S3 in Ihre Datenbanktabellen laden. Es gibt mehrere Möglichkeiten, Daten aus Amazon S3 zu laden.

- Sie können einen SQL-Client verwenden, um den SQL-Befehl CREATE TABLE auszuführen, um eine Tabelle in Ihrer Datenbank zu erstellen, und dann den Befehl SQL COPY verwenden, um Daten aus Amazon S3 zu laden. Der Amazon Redshift Query Editor v2 ist ein SQL-Client.
- Sie können den Ladeassistenten für den Amazon Redshift Query Editor v2 verwenden.

Dieses Tutorial zeigt zunächst, wie Sie den Amazon Redshift Query Editor v2 verwenden, um SQL-Befehle auszuführen, um Tabellen zu ERSTELLEN und Daten zu KOPIEREN. Starten Sie den Abfrage-Editor v2 über den Navigationsbereich der Amazon Redshift Redshift-Konsole. Stellen Sie im Query Editor v2 eine Verbindung zum `examplecluster` Cluster und zur Datenbank her, die `dev` mit Ihrem Admin-Benutzer `awsuser` benannt sind. Wählen Sie für dieses Tutorial Temporäre Anmeldeinformationen mit einem Datenbankbenutzernamen, wenn Sie die Verbindung herstellen. Einzelheiten zur Verwendung des Amazon Redshift Query Editors v2 finden Sie unter [Herstellen einer Verbindung zu einer Amazon Redshift-Datenbank](#) im Amazon Redshift Management Guide.



## Laden von Daten aus Amazon S3 mithilfe von SQL-Befehlen

Vergewissern Sie sich im Abfrage-Editor-Bereich des Abfrage-Editors v2, dass Sie mit dem `examplecluster` Cluster und der `dev` Datenbank verbunden sind. Erstellen Sie als Nächstes Tabellen in der Datenbank und laden Sie Daten in die Tabellen. In diesem Tutorial sind die Daten, die Sie laden, in einem Amazon S3 S3-Bucket verfügbar, auf den von vielen aus zugegriffen werden kann AWS-Regionen.

Das folgende Verfahren erstellt Tabellen und lädt Daten aus einem öffentlichen Amazon S3 S3-Bucket.

Verwenden Sie den Amazon Redshift Redshift-Abfrage-Editor v2, um die folgende Anweisung `create table` zu kopieren und auszuführen, um eine Tabelle im `public` Schema der `dev` Datenbank zu erstellen. Weitere Informationen zur Syntax finden Sie unter [CREATE TABLE](#) im Datenbankentwicklerhandbuch zu Amazon Redshift.

Um Daten mit einem SQL-Client wie dem Abfrage-Editor v2 zu erstellen und zu laden

1. Führen Sie den folgenden SQL-Befehl aus, um die `sales` Tabelle zu ERSTELLEN.

```
drop table if exists sales;
create table sales(
salesid integer not null,
listid integer not null distkey,
sellerid integer not null,
buyerid integer not null,
eventid integer not null,
dateid smallint not null sortkey,
qtysold smallint not null,
pricepaid decimal(8,2),
commission decimal(8,2),
saletime timestamp);
```

2. Führen Sie den folgenden SQL-Befehl aus, um die `date` Tabelle zu ERSTELLEN.

```
drop table if exists date;
create table date(
dateid smallint not null distkey sortkey,
caldate date not null,
day character(3) not null,
```



```
week smallint not null,  
month character(5) not null,  
qtr character(5) not null,  
year smallint not null,  
holiday boolean default('N'));
```

3. Laden Sie die sales Tabelle mit dem Befehl COPY aus Amazon S3.

#### Note

Wir empfehlen, den Befehl COPY zu verwenden, um große Datensätze von Amazon S3 in Amazon Redshift zu laden. Weitere Informationen zur COPY-Syntax finden Sie unter [COPY](#) im Datenbankentwicklerhandbuch zu Amazon Redshift.

Stellen Sie Authentifizierung für Ihren Cluster um Zugriff auf Amazon S3 in Ihrem Namen bereit, um die Beispieldaten zu laden. Sie stellen die Authentifizierung bereit, indem Sie auf die IAM-Rolle verweisen, die Sie erstellt und als default für Ihren Cluster festgelegt haben, als Sie bei der Erstellung des Clusters IAM-Rolle erstellen als Standard ausgewählt haben.

Laden Sie die sales Tabelle mit dem folgenden SQL-Befehl. Sie können optional die [Quelldaten für die sales Tabelle](#) von Amazon S3 herunterladen und anzeigen. .

```
COPY sales  
FROM 's3://redshift-downloads/tickit/sales_tab.txt'  
DELIMITER '\t'  
TIMEFORMAT 'MM/DD/YYYY HH:MI:SS'  
REGION 'us-east-1'  
IAM_ROLE default;
```

4. Laden Sie die date Tabelle mit dem folgenden SQL-Befehl. Sie können optional die [Quelldaten für die date Tabelle](#) von Amazon S3 herunterladen und anzeigen. .

```
COPY date  
FROM 's3://redshift-downloads/tickit/date2008_pipe.txt'  
DELIMITER '|'   
REGION 'us-east-1'  
IAM_ROLE default;
```

## Laden von Daten aus Amazon S3 mit dem Abfrage-Editor v2

Der Abfrage-Editor v2 vereinfacht das Laden von Daten bei Verwendung des Assistenten zum Laden von Daten. Der COPY-Befehl, der im Assistenten zum Laden von Daten aus Query Editor v2 generiert und verwendet wird, unterstützt viele der Parameter, die für die COPY-Befehlssyntax zum Laden von Daten aus Amazon S3 verfügbar sind. Weitere Informationen zum COPY-Befehl und zu seinen Optionen zum Kopieren und Laden aus Amazon S3 finden Sie unter [COPY aus dem Amazon Simple Storage Service](#) im Datenbankentwicklerhandbuch zu Amazon Redshift.

Um Ihre eigenen Daten aus Amazon S3 in Amazon Redshift zu laden, erfordert Amazon Redshift eine IAM-Rolle, die über die benötigten Berechtigungen zum Laden von Daten aus dem angegebenen Amazon S3 Bucket verfügt.

Öffnen Sie für dieses Tutorial zunächst den Amazon Redshift Query Editor v2 und stellen Sie eine Verbindung zu einer Datenbank her. Erstellen Sie als Nächstes die Tabelle, die die geladenen Daten enthalten soll. Laden Sie dann Ihre eigenen Daten aus Amazon S3 in Amazon Redshift. Weitere Informationen zur Arbeit mit dem Abfrage-Editor v2 finden Sie unter [Daten in eine Datenbank laden](#) im Amazon Redshift Management Guide.

## Erstellen Sie TICKIT-Daten in Ihrem Cluster

Sie können den vollständigen Satz von TICKIT-Tabellen erstellen und Daten auf folgende Weise in Ihren Cluster laden:

- Wenn Sie in der Amazon Redshift Redshift-Konsole einen Cluster erstellen, haben Sie die Möglichkeit, TICKIT-Beispieldaten gleichzeitig zu laden. Wählen Sie in der Amazon Redshift Redshift-Konsole Clusters, Create cluster aus. Wählen Sie im Abschnitt Beispieldaten die Option Beispieldaten laden aus. Amazon Redshift lädt seinen Beispieldatensatz während der Clustererstellung automatisch in Ihre Amazon Redshift dev Redshift-Cluster-Datenbank.
- Mit dem Amazon Redshift Query Editor v2 können Sie TICKIT-Daten in eine Beispieldatenbank mit dem Namen `sample_data_dev` laden. Navigieren Sie zur Datenbank `sample_data_dev`, klicken Sie dann auf Tickit und wählen Sie im Kontextmenü Beispielnotizbücher öffnen. Der Amazon Redshift Query Editor v2 erstellt die Beispieldatenbank zusammen mit einem Beispielnotizbuch namens `tickit-sample-notebook`. Sie können dieses Notizbuch ausführen, um Daten in der Beispieldatenbank abzufragen.

Einzelheiten zu den TICKIT-Daten finden Sie unter [Beispieldatenbank](#) im Amazon Redshift Database Developer Guide.

## Schritt 5: Beispielabfragen mit dem Abfrage-Editor testen

Informationen zum Einrichten und Verwenden des Amazon Redshift-Abfrage-Editors v2 zum Abfragen einer Datenbank finden Sie unter [Arbeiten mit dem Abfrage-Editor v2](#) im Amazon Redshift Management Guide.

Testen Sie jetzt einige Beispielabfragen wie folgt. Weitere Informationen zur Arbeit mit dem SELECT-Befehl finden Sie unter [SELECT](#) im Amazon Redshift Database Developer Guide.

```
-- Get definition for the sales table.
SELECT *
FROM pg_table_def
WHERE tablename = 'sales';
```

```
-- Find total sales on a given calendar date.
SELECT sum(qtysold)
FROM   sales, date
WHERE  sales.dateid = date.dateid
AND    caldate = '2008-01-05';
```

```
-- Find top 10 buyers by quantity.
SELECT firstname, lastname, total_quantity
FROM   (SELECT buyerid, sum(qtysold) total_quantity
        FROM   sales
        GROUP BY buyerid
        ORDER BY total_quantity desc limit 10) Q, users
WHERE  Q.buyerid = userid
ORDER BY Q.total_quantity desc;
```

```
-- Find events in the 99.9 percentile in terms of all time gross sales.
SELECT eventname, total_price
FROM   (SELECT eventid, total_price, ntile(1000) over(order by total_price desc) as
        percentile
        FROM (SELECT eventid, sum(pricepaid) total_price
              FROM   sales
              GROUP BY eventid)) Q, event E
WHERE  Q.eventid = E.eventid
AND    percentile = 1
ORDER BY total_price desc;
```

## Schritt 6: Umgebung zurücksetzen

In den vorherigen Schritten haben Sie erfolgreich einen Amazon Redshift-Cluster erstellt, Daten in Tabellen geladen und Daten mit einem SQL-Client wie dem Amazon Redshift Query Editor v2 abgefragt.

Wenn Sie dieses Tutorial abgeschlossen haben, empfehlen wir, dass Sie Ihre Umgebung auf den vorherigen Zustand zurücksetzen, indem Sie Ihren Beispielcluster löschen. Es fallen so lange Amazon-Redshift-Nutzungsgebühren, bis Sie den Cluster löschen.

Möglicherweise möchten Sie den Beispielcluster jedoch weiterlaufen lassen, wenn Sie Aufgaben in anderen Amazon Redshift Redshift-Handbüchern oder Aufgaben ausprobieren möchten, die unter beschrieben sind. [Häufige Datenbankaufgaben](#)

### Löschen eines Clusters

1. Melden Sie sich bei der Amazon Redshift Redshift-Konsole an AWS Management Console und öffnen Sie sie unter <https://console.aws.amazon.com/redshiftv2/>.
2. Wählen Sie im Navigationsmenü Clusters (Cluster) aus, um Ihre Liste der Cluster anzuzeigen.
3. Wählen Sie den Cluster `examplecluster` aus. Klicken Sie bei Actions auf Delete. Die Seite Delete cluster (Cluster löschen) wird angezeigt.
4. Bestätigen Sie, dass der Cluster gelöscht werden soll. Wählen Sie dann Delete cluster (Cluster löschen) aus.

Auf der Seite mit der Clusterliste wird der Clusterstatus aktualisiert, wenn der Cluster gelöscht wird.

Nach Abschluss dieses Tutorials finden Sie weitere Informationen über Amazon Redshift sowie die nächsten Schritte unter [Weitere Ressourcen](#).

# Häufige Datenbankaufgaben

Sowohl Redshift Serverless Data Warehouses als auch Amazon Redshift Cluster Data Warehouses enthalten Datenbanken. Nachdem Sie Ihr Data Warehouse gestartet haben, können Sie die meisten Datenbankaktionen mithilfe von SQL-Befehlen verwalten. Mit wenigen Ausnahmen sind die Funktionalität und Syntax von SQL für alle Amazon Redshift Redshift-Datenbanken identisch. Einzelheiten zu den in Amazon Redshift verfügbaren SQL-Befehlen finden Sie unter [SQL-Befehle](#) im Amazon Redshift Database Developer Guide.

Wenn Sie Ihr Data Warehouse erstellen, erstellt Amazon Redshift in den meisten Szenarien auch die dev Standarddatenbank. Nachdem Sie eine Verbindung mit der dev Datenbank hergestellt haben, können Sie eine weitere Datenbank erstellen.

In den folgenden Schritten werden allgemeine Datenbankaufgaben bei der Arbeit mit Amazon Redshift Redshift-Datenbanken beschrieben. Die Schritte beginnen mit dem Erstellen einer Datenbank. Wenn Sie mit dem letzten Schritt fortfahren, können Sie alle von Ihnen erstellten Ressourcen löschen, indem Sie die Datenbank löschen.

Die Beispiele in diesem Abschnitt setzen Folgendes voraus:

- Sie haben ein Amazon Redshift Data Warehouse erstellt.
- Sie haben über Ihr SQL-Client-Tool, z. B. den Amazon Redshift Query Editor v2, eine Verbindung zum Data Warehouse hergestellt. Weitere Informationen zum Abfrage-Editor v2 finden Sie unter [Abfragen einer Datenbank mit dem Amazon Redshift-Abfrage-Editor v2](#) im Amazon Redshift Management Guide.

## Verbindung zu Amazon Redshift Data Warehouses herstellen

Um eine Verbindung zu Amazon Redshift Redshift-Clustern herzustellen, erweitern Sie auf der Clusterseite der Amazon Redshift Redshift-Konsole die Option Mit Amazon Redshift Redshift-Clustern Connect und führen Sie einen der folgenden Schritte aus:

- Wählen Sie Daten abfragen, um mit dem Abfrage-Editor v2 Abfragen in Datenbanken auszuführen, die von Ihrem Amazon Redshift Redshift-Cluster gehostet werden. Nachdem Sie Ihren Cluster erstellt haben, können Sie mit dem Abfrageeditor v2 sofort Abfragen ausführen.

Weitere Informationen finden Sie unter [Abfragen einer Datenbank mit dem Amazon Redshift Query Editor v2](#) im Amazon Redshift Management Guide.

- Wählen Sie unter Arbeiten mit Ihren Client-Tools Ihren Cluster aus und stellen Sie über Ihre Client-Tools mithilfe von JDBC- oder ODBC-Treibern eine Verbindung zu Amazon Redshift her, indem Sie die URL des JDBC- oder ODBC-Treibers kopieren. Verwenden Sie diese URL von Ihrem Client-Computer oder Ihrer Instance aus. Schreiben Sie Ihre Anwendungen so, dass sie JDBC- oder ODBC-API-Operationen für den Zugriff auf Daten nutzen, oder verwenden Sie SQL-Client-Tools, die JDBC oder ODBC unterstützen.

Weitere Informationen darüber, wie Sie Ihre Cluster-Verbindungszeichenfolge finden, erhalten Sie unter [Suche der Zeichenfolge für die Verbindung mit dem Cluster](#).

- Wenn Ihr SQL-Client-Tool einen Treiber benötigt, können Sie Ihren JDBC- oder ODBC-Treiber auswählen, um einen betriebssystemspezifischen Treiber für die Verbindung mit Amazon Redshift von Ihren Client-Tools herunterzuladen.

Weitere Informationen zum Installieren des geeigneten Treibers für Ihren SQL-Client finden Sie unter [Konfigurieren einer Verbindung mit JDBC-Treiberversion 2.0](#).

Weitere Informationen zum Konfigurieren einer ODBC-Verbindung finden Sie unter [Konfigurierung einer ODBC-Verbindung](#).

Gehen Sie auf der Serverless-Dashboard-Seite der Amazon Redshift-Konsole wie folgt vor, um eine Verbindung zum Redshift Serverless Data Warehouse herzustellen:

- Verwenden Sie den Amazon Redshift Query Editor v2, um Abfragen in Datenbanken auszuführen, die von Ihrem Redshift Serverless Data Warehouse gehostet werden. Nachdem Sie Ihr Data Warehouse erstellt haben, können Sie Abfragen sofort mit dem Abfrage-Editor v2 ausführen.

Weitere Informationen finden Sie unter [Abfragen einer Datenbank mit dem Abfrage-Editor v2 von Amazon Redshift](#).

- Stellen Sie mithilfe von JDBC- oder ODBC-Treibern eine Verbindung von Ihren Client-Tools zu Amazon Redshift her, indem Sie die URL des JDBC- oder ODBC-Treibers kopieren.

Um mit Daten in Ihrem Data Warehouse arbeiten zu können, benötigen Sie JDBC- oder ODBC-Treiber für die Konnektivität von Ihrem Client-Computer oder Ihrer Instanz aus. Schreiben Sie Ihre Anwendungen so, dass sie JDBC- oder ODBC-API-Operationen für den Zugriff auf Daten nutzen, oder verwenden Sie SQL-Client-Tools, die JDBC oder ODBC unterstützen.

Weitere Informationen zum Auffinden Ihrer Verbindungszeichenfolge finden Sie unter [Connecting to Redshift Serverless](#) im Amazon Redshift Management Guide.

## Schritt 1: Erstellen einer Datenbank

Nachdem Sie sich vergewissert haben, dass Ihr Data Warehouse betriebsbereit ist, können Sie eine Datenbank erstellen. In dieser Datenbank erstellen Sie Tabellen, laden Daten und führen Abfragen aus. Ein Data Warehouse kann mehrere Datenbanken hosten. Sie können beispielsweise eine Datenbank für Verkaufsdaten SALESDB und eine Datenbank für Bestelldaten ORDERSDB in demselben Data Warehouse benennen.

Um eine Datenbank mit dem Namen zu erstellen **SALESDB**, führen Sie den folgenden Befehl in Ihrem SQL-Client-Tool aus.

```
CREATE DATABASE salesdb;
```

### Note

Stellen Sie nach der Ausführung des Befehls sicher, dass Sie die Liste der Objekte in Ihrem Data Warehouse in Ihrem SQL-Client-Tool aktualisieren, um die neuen Objekte zu sehen `salesdb`.

Für diese Übung übernehmen wir die Standardeinstellungen. Weitere Informationen über Befehlsoptionen finden Sie unter [CREATE DATABASE](#) im Datenbankentwicklerhandbuch zu Amazon Redshift. Informationen zum Löschen einer Datenbank und ihres Inhalts finden Sie unter [DROP DATABASE](#) im Amazon Redshift Database Developer Guide.

Nachdem Sie die SALESDB-Datenbank erstellt haben, können Sie von Ihrem SQL-Client aus eine Verbindung zu der neuen Datenbank herstellen. Verwenden Sie dieselben Verbindungsparameter wie für Ihre aktuelle Verbindung, ändern Sie aber den Namen der Datenbank in SALESDB.

## Schritt 2: Erstellen Sie einen Benutzer

Standardmäßig hat nur der Admin-Benutzer, den Sie beim Start des Data Warehouse erstellt haben, Zugriff auf die Standarddatenbank im Data Warehouse. Um anderen Benutzern den Zugriff zu gewähren, erstellen Sie ein oder mehrere Benutzerkonten. Datenbankbenutzerkonten gelten global für alle Datenbanken in einem Data Warehouse und nicht für jede einzelne Datenbank.

Verwenden Sie den Befehl `CREATE USER`, um einen neuen Benutzer zu erstellen. Wenn Sie einen neuen Benutzer erstellen, geben Sie dessen Namen und ein Passwort an. Es wird empfohlen, dass

Sie ein Passwort für den Benutzer angeben. Dies muss aus 8–64 Zeichen bestehen und mindestens einen Großbuchstaben, einen Kleinbuchstaben und eine Ziffer enthalten.

Führen Sie beispielsweise zur Erstellung eines Benutzers mit dem Namen **GUEST** und dem Passwort **ABCd4321** den folgenden Befehl aus:

```
CREATE USER GUEST PASSWORD 'ABCd4321';
```

Um eine Verbindung mit der SALESDB-Datenbank als der GUEST-Benutzer herzustellen, verwenden Sie das Passwort, das Sie bei der Benutzererstellung gewählt haben, zum Beispiel ABCd4321.

Informationen über weitere Befehlsoptionen finden Sie unter [CREATE USER](#) im Datenbankentwicklerhandbuch zu Amazon Redshift.

## Schritt 3: Erstellen Sie ein Schema

Nach dem Erstellen einer neuen Datenbank können Sie ein neues Schema in der aktuellen Datenbank erstellen. Ein Schema ist ein Namespace, der benannte Datenbankobjekte wie Tabellen, Ansichten und benutzerdefinierte Funktionen (User-Defined Functions, UDFs) enthält. Eine Datenbank kann ein oder mehrere Schemata enthalten, und jedes Schema gehört nur zu einer Datenbank. Zwei Schemata können verschiedene Objekte mit demselben Namen haben.

Sie können mehrere Schemata in derselben Datenbank erstellen, um Daten so zu organisieren, wie Sie möchten, oder um Ihre Daten funktional zu gruppieren. Sie können beispielsweise ein Schema erstellen, um alle Staging-Daten zu speichern, und ein anderes Schema zum Speichern aller Berichtstabellen. Sie können auch verschiedene Schemata erstellen, um Daten zu speichern, die für verschiedene Unternehmensgruppen relevant sind, die sich in derselben Datenbank befinden. Jedes Schema kann verschiedene Datenbankobjekte speichern, wie etwa Tabellen, Ansichten und benutzerdefinierte Funktionen (User-Defined Functions, UDFs). Darüber hinaus können Sie Schemata mit der AUTHORIZATION-Klausel erstellen. Diese Klausel gewährt einem bestimmten angegebenen Benutzer Besitz oder legt ein Kontingent für den maximalen Speicherplatz fest, den das angegebene Schema verwenden kann.

Amazon Redshift erstellt automatisch ein Schema namens `public` für jede neue Datenbank. Wenn Sie den Schemanamen beim Erstellen von Datenbankobjekten nicht angeben, gehen die Objekte in das `public`-Schema über.

Um auf ein Objekt in einem Schema zuzugreifen, qualifizieren Sie das Objekt mithilfe der `schema_name.table_name`-Notation. Der qualifizierte Name des Schemas besteht aus dem



Schemanamen und dem Tabellennamen, die durch einen Punkt getrennt sind. Zum Beispiel kann ein sales-Schema eine price-Tabelle und ein inventory-Schema eine price-Tabelle haben. Wenn Sie die price-Tabelle referenzieren, müssen Sie sie als sales.price oder inventory.price qualifizieren.

Im folgenden Beispiel wird ein Schema mit dem Namen **SALES** für den Benutzer GUEST erstellt.

```
CREATE SCHEMA SALES AUTHORIZATION GUEST;
```

Weitere Informationen über Befehlsoptionen finden Sie unter [CREATE SCHEMA](#) im Datenbankentwicklerhandbuch zu Amazon Redshift.

Führen Sie den folgenden Befehl aus, um die Liste der Schemata in Ihrer Datenbank anzuzeigen.

```
select * from pg_namespace;
```

Die Ausgabe sollte in etwa folgendermaßen aussehen:

nspname	nspowner	nspacl
sales	100	
pg_toast	1	
pg_internal	1	
catalog_history	1	
pg_temp_1	1	
pg_catalog	1	{rdsdb=UC/rdsdb,=U/rdsdb}
public	1	{rdsdb=UC/rdsdb,=U/rdsdb}
information_schema	1	{rdsdb=UC/rdsdb,=U/rdsdb}

Weitere Informationen zum Abfragen von Katalogtabellen finden Sie unter [Abfragen der Katalogtabellen](#) im Datenbankentwicklerhandbuch zu Amazon Redshift.

Verwenden Sie die GRANT-Anweisung, um Benutzern Berechtigungen für die Schemata zu erteilen.

Im folgenden Beispiel wird dem GUEST Benutzer die Berechtigung erteilt, mithilfe einer SELECT-Anweisung Daten aus allen Tabellen oder Ansichten im SALES Schema auszuwählen.

```
GRANT SELECT ON ALL TABLES IN SCHEMA SALES TO GUEST;
```

Im folgenden Beispiel werden dem GUEST Benutzer alle verfügbaren Berechtigungen gleichzeitig gewährt.

```
GRANT ALL ON SCHEMA SALES TO GUEST;
```

## Schritt 4: Erstellen Sie eine Tabelle

Nach dem Erstellen Ihrer neuen Datenbank erstellen Sie Tabellen für Ihre Daten. Geben Sie die Spalteninformationen an, wenn Sie die Tabelle erstellen.

Zum Beispiel können Sie mit dem folgenden Befehl eine Tabelle namens **DEMO** erstellen.

```
CREATE TABLE Demo (  
    PersonID int,  
    City varchar (255)  
);
```

Standardmäßig werden neue Datenbankobjekte, wie z. B. Tabellen, in dem Standardschema mit dem Namen erstellt, das bei der Data Warehouse-Erstellung `public` erstellt wurde, erstellt. Sie können ein anderes Schema verwenden, um Datenbankobjekte zu erstellen. Weitere Informationen über Schemata finden Sie unter [Verwalten der Datenbanksicherheit](#) im Datenbankentwicklerhandbuch zu Amazon Redshift.

Darüber hinaus können Sie mit der `schema_name.object_name`-Notation auch eine Tabelle im SALES-Schema erstellen.

```
CREATE TABLE SALES.DEMO (  
    PersonID int,  
    City varchar (255)  
);
```

Um Schemas und ihre Tabellen anzuzeigen und zu überprüfen, können Sie den Amazon Redshift Redshift-Abfrage-Editor v2 verwenden. Oder Sie können die Liste der Tabellen in Schemata mithilfe von Systemansichten ansehen. Weitere Informationen finden Sie unter [Schritt 6: Fragen Sie die Systemtabellen ab](#).

Die Spalten `encoding`, `distkey` und `sortkey` werden von Amazon Redshift für die parallele Verarbeitung verwendet. Für weitere Informationen zum Entwurf von Tabellen mit diesen Elementen siehe [Bewährte Methoden für die Gestaltung von Tabellen mit Amazon Redshift](#).

## Einfügen von Datenzeilen in eine Tabelle

Nach der Erstellung der Tabelle fügen Sie Datenzeilen darin ein.

### Note

Zeilen werden mit dem Befehl [INSERT](#) in Tabellen eingefügt. Verwenden Sie für Standard-Masseneinfügungen den Befehl [COPY](#). Weitere Informationen finden Sie unter [Verwenden eines COPY-Befehls zum Laden von Daten](#).

Um zum Beispiel Werte in die Tabelle DEMO einzufügen, führen Sie folgenden Befehl aus.

```
INSERT INTO DEMO VALUES (781, 'San Jose'), (990, 'Palo Alto');
```

Um Daten in eine Tabelle einzufügen, die sich in einem bestimmten Schema befindet, führen Sie den folgenden Befehl aus.

```
INSERT INTO SALES.DEMO VALUES (781, 'San Jose'), (990, 'Palo Alto');
```

## Auswahl von Daten aus einer Tabelle

Nachdem Sie eine Tabelle erstellt und mit Daten gefüllt haben, verwenden Sie eine SELECT-Anweisung, um die in der Tabelle enthaltenen Daten anzuzeigen. Die Anweisung SELECT \* gibt alle Spaltennamen und Zeilenwerte für alle Daten in einer Tabelle zurück. Die Verwendung von SELECT ist eine gute Möglichkeit, um zu prüfen, ob kürzlich hinzugefügte Daten korrekt in die Tabelle eingefügt wurden.

Um die Daten anzuzeigen, die Sie in die Tabelle **DEMO** eingegeben haben, führen Sie den folgenden Befehl aus:

```
SELECT * from DEMO;
```

Das Ergebnis sollte wie das folgende aussehen.

```
personid | city
-----+-----
       781 | San Jose
```

```
990 | Palo Alto
(2 rows)
```

Weitere Informationen zur Verwendung der SELECT-Anweisung zur Abfrage von Tabellen finden Sie unter [SELECT](#).

## Schritt 5: Daten laden

Viele der Beispiele in diesem Handbuch verwenden den TICKIT-Beispieldatensatz. Sie können die Datei [ticketdb.zip](#) herunterladen. Diese enthält einzelne Beispieldatendateien. Anschließend können Sie die Beispieldaten in Ihren eigenen Amazon S3 S3-Bucket hochladen.

Um die Beispieldaten für Ihre Datenbank zu laden, erstellen Sie zuerst die Tabellen. Verwenden Sie dann den COPY-Befehl, um die Tabellen mit Beispieldaten zu laden, die in einem Amazon S3 Bucket gespeichert sind. Weitere Informationen zu den Schritten für die Erstellung von Tabellen und das Laden von Beispieldaten finden Sie unter [Schritt 4: Daten aus Amazon S3 in Amazon Redshift laden](#).

## Schritt 6: Fragen Sie die Systemtabellen ab

Zusätzlich zu den Tabellen, die Sie erstellen, enthält Ihr Data Warehouse eine Reihe von Systemtabellen und Ansichten. Diese Systemtabellen enthalten Informationen zu Ihrer Installation sowie über die verschiedenen Abfragen und Prozesse, die auf dem System ausgeführt werden. Sie können diese Systemtabellen abfragen, um Informationen zu Ihrer Datenbank zu erhalten. Weitere Informationen finden Sie unter [Referenz zu Systemtabellen und -ansichten](#) im Amazon Redshift Database Developer Guide. Die Beschreibung jeder Tabelle oder Ansicht gibt an, ob eine Tabelle für alle Benutzer oder nur für Superuser sichtbar ist. Um nur für Superuser sichtbare Tabellen anzuzeigen, melden Sie sich als Superuser an.

## Anzeigen einer Liste von Tabellennamen

Um eine Liste aller Tabellen in einem Schema anzuzeigen, können Sie die Systemkatalogtabelle PG\_TABLE\_DEF abfragen. Sie können zunächst die Einstellung für prüfen `search_path`.

```
SHOW search_path;
```

Das Ergebnis sollte in etwa wie folgt aussehen.

```
search_path
```

```
-----
$user, public
```

Im folgenden Beispiel wird das SALES-Schema dem Suchpfad hinzugefügt und es werden alle Tabellen im SALES-Schema angezeigt.

```
set search_path to '$user', 'public', 'sales';
```

```
SHOW search_path;
```

```
search_path
```

```
-----
"$user", public, sales
```

```
select * from pg_table_def where schemaname = 'sales';
```

schemaname	tablename	column	type	encoding	distkey	sortkey	notnull
sales	demo	personid	integer	az64	f		
							0   f
sales	demo	city	character varying(255)	lzo	f		
							0   f

Im folgenden Beispiel wird eine Liste aller Tabellen mit dem Namen DEMO in allen Schemata der aktuellen Datenbank angezeigt.

```
set search_path to '$user', 'public', 'sales';
```

```
select * from pg_table_def where tablename = 'demo';
```

schemaname	tablename	column	type	encoding	distkey	sortkey	notnull
public	demo	personid	integer	az64	f		
							0   f
public	demo	city	character varying(255)	lzo	f		
							0   f
sales	demo	personid	integer	az64	f		
							0   f

```
sales      | demo      | city      | character varying(255) | lzo      | f      |
0 | f
```

Weitere Informationen finden Sie in der Tabelle [PG\\_TABLE\\_DEF](#).

Sie können auch den Amazon Redshift Redshift-Abfrage-Editor v2 verwenden, um alle Tabellen in einem bestimmten Schema anzuzeigen, indem Sie zunächst eine Datenbank auswählen, zu der Sie eine Verbindung herstellen möchten.

## Anzeigen von Benutzern

Sie können den Katalog PG\_USER abfragen, um eine Liste aller Benutzer zusammen mit Benutzer-ID (USESYSID) und Benutzerberechtigungen anzuzeigen.

```
SELECT * FROM pg_user;
```

```
username | usesysid | usecreatedb | usesuper | usecatupd | passwd | valuntil |
useconfig
-----+-----+-----+-----+-----+-----+-----
+-----
rdsdb    |         1 | true        | true     | true      | ***** | infinity |
awsuser  |        100 | true        | true     | false     | ***** |          |
guest    |        104 | true        | false    | false     | ***** |          |
```

Der Benutzername `rdsdb` wird intern von Amazon Redshift für Routine-Verwaltungs- und Wartungsaufgaben verwendet. Sie können Ihre Abfrage so filtern, dass nur benutzerdefinierte Benutzernamen angezeigt werden, indem Sie Ihrer SELECT-Anweisung `where usesysid > 1` hinzufügen.

```
SELECT * FROM pg_user WHERE usesysid > 1;
```

```
username | usesysid | usecreatedb | usesuper | usecatupd | passwd | valuntil |
useconfig
-----+-----+-----+-----+-----+-----+-----
+-----
awsuser  |        100 | true        | true     | false     | ***** |          |
guest    |        104 | true        | false    | false     | ***** |          |
```

## Anzeigen aktueller Abfragen

Im vorherigen Beispiel ist die Benutzer-ID (`user_id`) für `adminuser` 100. Um die vier zuletzt von ausgeführten Abfragen aufzulisten `adminuser`, können Sie die Ansicht `SYS_QUERY_HISTORY` abfragen.

Sie können diese Ansicht verwenden, um die Abfrage-ID (`query_id`) oder die Prozess-ID (`session_id`) für eine kürzlich ausgeführte Abfrage zu finden. Sie können diese Ansicht auch verwenden, um zu überprüfen, wie lange eine Abfrage in Anspruch nahm. `SYS_QUERY_HISTORY` enthält die ersten 4.000 Zeichen der Abfragezeichenfolge (`query_text`), um Ihnen das Auffinden einer bestimmten Abfrage zu erleichtern. Verwenden Sie die `LIMIT`-Klausel mit Ihrer `SELECT`-Anweisung, um die Ergebnisse einzuschränken.

```
SELECT query_id, session_id, elapsed_time, query_text
FROM sys_query_history
WHERE user_id = 100
ORDER BY start_time desc
LIMIT 4;
```

Das Ergebnis sieht in etwa wie folgt aus.

query_id	session_id	elapsed_time	query_text
892	21046	55868	SELECT query, pid, elapsed, substring from ...
620	17635	1296265	SELECT query, pid, elapsed, substring from ...
610	17607	82555	SELECT * from DEMO;
596	16762	226372	INSERT INTO DEMO VALUES (100);

## Ermitteln Sie die Sitzungs-ID einer laufenden Abfrage

Um Systemtabelleninformationen zu einer Abfrage abzurufen, müssen Sie möglicherweise die Sitzungs-ID (Prozess-ID) angeben, die dieser Abfrage zugeordnet ist. Oder Sie müssen möglicherweise die Sitzungs-ID für eine Abfrage ermitteln, die noch ausgeführt wird. Beispielsweise benötigen Sie die Sitzungs-ID, wenn Sie eine Abfrage abbrechen müssen, deren Ausführung auf einem bereitgestellten Cluster zu lange dauert. Sie können die `STV_RECENTS`-Systemtabelle abfragen, um eine Liste von Sitzungs-IDs für laufende Abfragen zusammen mit der entsprechenden

Abfragezeichenfolge zu erhalten. Wenn Ihre Abfrage mehrere Sitzungen zurückgibt, können Sie anhand des Abfragetextes ermitteln, welche Sitzungs-ID Sie benötigen.

Führen Sie die folgende SELECT-Anweisung aus, um die Sitzungs-ID einer laufenden Abfrage zu ermitteln.

```
SELECT session_id, user_id, start_time, query_text
FROM sys_query_history
WHERE status='running';
```

## Schritt 7: Stornieren Sie eine Abfrage

Wenn Sie eine Abfrage ausführen, die zu lange dauert oder zu viele Ressourcen verbraucht, brechen Sie die Abfrage ab. Zum Beispiel: Erstellen Sie eine Liste von Ticketverkäufern, die die Namen der Verkäufer und die Anzahl der verkauften Tickets enthält. Die folgende Abfrage wählt Daten aus der SALES-Tabelle und der USERS-Tabelle aus und verbindet beide Tabellen durch den Abgleich von SELLERID und USERID in der WHERE-Klausel.

```
SELECT sellerid, firstname, lastname, sum(qtysold)
FROM sales, users
WHERE sales.sellerid = users.userid
GROUP BY sellerid, firstname, lastname
ORDER BY 4 desc;
```

Das Ergebnis sieht in etwa wie folgt aus.

sellerid	firstname	lastname	sum
48950	Nayda	Hood	184
19123	Scott	Simmons	164
20029	Drew	Mcguire	164
36791	Emerson	Delacruz	160
13567	Imani	Adams	156
9697	Dorian	Ray	156
41579	Harrison	Durham	156
15591	Phyllis	Clay	152
3008	Lucas	Stanley	148
44956	Rachel	Villarreal	148



**Note**

Dies ist eine komplexe Abfrage. Für dieses Tutorial müssen Sie sich über den Aufbau dieser Abfrage keine Gedanken machen.

Die vorherige Abfrage dauert wenige Sekunden und gibt 2 102 Zeilen aus.

Angenommen, Sie hätten die WHERE-Klausel vergessen.

```
SELECT sellerid, firstname, lastname, sum(qtysold)
FROM sales, users
GROUP BY sellerid, firstname, lastname
ORDER BY 4 desc;
```

Der Ergebnissatz enthält dann alle Zeilen in der SALES-Tabelle, multipliziert mit allen Zeilen in der USERS-Tabelle (49989\*3766). Dies ist eine so genannte Cartesische Verbindung, die nicht zu empfehlen ist. Das Ergebnis sind mehr als 188 Millionen Zeilen, und die Verarbeitungszeit ist extrem lang.

Um eine laufende Abfrage abubrechen, verwenden Sie den Befehl CANCEL mit der Sitzungs-ID der Abfrage. Mit dem Amazon Redshift Redshift-Abfrage-Editor v2 können Sie eine Abfrage abbrechen, indem Sie auf die Schaltfläche Abbrechen klicken, während die Abfrage ausgeführt wird.

Um die Sitzungs-ID zu finden, starten Sie eine neue Sitzung und fragen Sie die Tabelle STV\_RECENTS ab, wie im vorherigen Schritt gezeigt. Das folgende Beispiel zeigt, wie Sie die Ergebnisse lesbarer machen können. Verwenden Sie dazu die TRIM-Funktion, um nachfolgende Leerzeichen abzuschneiden, und zeigen Sie nur die ersten 20 Zeichen der Abfragezeichenfolge an.

Führen Sie die folgende SELECT-Anweisung aus, um die Sitzungs-ID einer laufenden Abfrage zu ermitteln.

```
SELECT user_id, session_id, start_time, query_text
FROM sys_query_history
WHERE status='running';
```

Das Ergebnis sieht in etwa wie folgt aus.

user_id	session_id	start_time	query_text
---------	------------	------------	------------

```
-----+-----+-----  
+-----  
100      |      1073791534 | 2024-03-19 22:26:21.205739 | SELECT user_id, session_id,  
start_time, query_text FROM ...
```

Führen Sie den folgenden Befehl aus `1073791534`, um die Abfrage mit der Sitzungs-ID abzurechnen.

```
CANCEL 1073791534;
```

### Note

Der Befehl `CANCEL` stoppt eine Transaktion nicht. Um eine Transaktion zu stoppen oder rückgängig zu machen, müssen Sie den Befehl `ABORT` oder `ROLLBACK` verwenden. Um eine mit einer Transaktion verbundene Abfrage abzurechnen, brechen Sie zuerst die Abfrage ab und stoppen Sie dann die Transaktion.

Wenn die abgebrochene Abfrage mit einer Transaktion verbunden ist, verwenden Sie den Befehl `ABORT` oder `ROLLBACK`, um die Transaktion abzurechnen und alle an den Daten vorgenommenen Änderungen zu verwerfen:

```
ABORT;
```

Sie können nur Ihre eigenen Abfragen abrechnen, sofern Sie nicht als Superuser angemeldet sind. Superuser können alle Abfragen abrechnen.

Wenn Ihr Abfragetool nicht die gleichzeitige Ausführung von Abfragen unterstützt, starten Sie zum Abbruch der Abfrage eine weitere Sitzung.

Weitere Informationen zum Stornieren einer Abfrage finden Sie unter [CANCEL](#) im Amazon Redshift Database Developer Guide.

## Abbrechen einer Abfrage mit der Superuser-Warteschlange

Wenn in Ihrer aktuellen Sitzung zu viele Abfragen gleichzeitig ausgeführt werden, können Sie möglicherweise erst dann den `CANCEL`-Befehl ausführen, wenn eine andere Abfrage abgeschlossen ist. Führen Sie in diesem Fall den `CANCEL`-Befehl mit einer anderen Workload-Verwaltungs-Abfragewarteschlange aus.

Workload-Verwaltung ermöglicht Ihnen die Ausführung von Abfragen in verschiedenen Abfragewarteschlangen, so dass Sie nicht warten müssen, bis eine andere Abfrage abgeschlossen ist. Der Workload Manager erstellt eine separate Warteschlange mit der Bezeichnung „Superuser-Warteschlange“, die Sie für Fehlerbehebungsziecke verwenden können. Um die Superuser-Warteschlange verwenden zu können, melden Sie sich als Superuser an und setzen Sie die Abfragegruppe mit dem SET-Befehl auf „Superuser“. Setzen Sie nach der Ausführung Ihrer Befehle die Abfragegruppe mit dem RESET-Befehl wieder zurück.

Um eine Abfrage mithilfe der Superuser-Warteschlange abubrechen, führen Sie diese Befehle aus.

```
SET query_group TO 'superuser';  
CANCEL 1073791534;  
RESET query_group;
```

# Abfragen von Datenquellen außerhalb Ihrer Amazon-Redshift-Datenbank

Im Folgenden erfahren Sie mehr über die ersten Schritte beim Abfragen von Daten in Remote-Quellen, einschließlich Amazon-Redshift-Remote-Clustern. Sie erhalten auch Informationen zum Trainieren von Machine Learning (ML)-Modellen mit Amazon Redshift.

## Themen

- [Abfragen Ihres Data Lake](#)
- [Abfragen von Daten in Remote-Datenquellen](#)
- [Zugriff auf Daten in anderen Amazon-Redshift-Clustern](#)
- [Training von Machine-Learning-Modellen mit Amazon-Redshift-Daten](#)

## Abfragen Ihres Data Lake

Mit Amazon Redshift Spectrum können Sie Daten in Amazon-S3-Dateien abfragen, ohne die Daten in Amazon-Redshift-Tabellen laden zu müssen. Amazon Redshift bietet SQL-Funktionen für die schnelle Online-Analyseverarbeitung (OLAP) von sehr großen Datensätzen, die sowohl in Amazon-Redshift-Clustern als auch Amazon-S3-Data-Lakes gespeichert sind. Sie können Daten in vielen Formaten abfragen, darunter Parquet, ORC, RCFile, TextFile, SequenceFile, RegexSerde, OpenCSV und AVRO. Um die Struktur der Dateien in Amazon S3 zu definieren, erstellen Sie externe Schemata und Tabellen. Anschließend verwenden Sie einen externen Datenkatalog wie AWS Glue oder Ihren eigenen Apache-Hive-Metastore. Änderungen an einem der Datenkatalogtypen sind sofort für jeden Ihrer Amazon-Redshift-Cluster verfügbar.

Nachdem Ihre Daten bei einem AWS Glue-Datenkatalog registriert und mit AWS Lake Formation aktiviert wurden, können Sie sie mit Redshift Spectrum abfragen.

Redshift Spectrum befindet sich auf dedizierten Amazon-Redshift-Servern, die von Ihrem Cluster unabhängig sind. Redshift Spectrum verschiebt viele datenverarbeitungsintensive Aufgaben, wie etwa die Prädikatfilterung und -aggregation, auf die Redshift-Spectrum-Ebene. Redshift Spectrum lässt sich auch intelligent skalieren, um die Vorteile der massiv parallelen Verarbeitung zu nutzen.

Sie können die externen Tabellen in einer oder mehreren Spalten partitionieren, um die Abfrageleistung durch Partitionseliminierung zu optimieren. Sie können die externen Tabellen mit

Amazon-Redshift-Tabellen abfragen und verknüpfen. Sie können externe Tabellen von mehreren Amazon-Redshift-Clustern abrufen und die Amazon-S3-Daten aus jedem Cluster in derselben AWS-Region abfragen. Wenn Sie Amazon-S3-Datendateien aktualisieren, stehen diese Daten sofort zur Abfrage von allen Ihren Amazon-Redshift-Clustern aus zur Verfügung.

Weitere Informationen zu Redshift Spectrum, einschließlich zur Arbeit mit Redshift Spectrum und Data Lakes, finden Sie unter [Erste Schritte mit Amazon Redshift Spectrum](#) im Datenbankentwicklerhandbuch zu Amazon Redshift.

## Abfragen von Daten in Remote-Datenquellen

Sie können Daten mithilfe einer Verbundabfrage aus einer Amazon-RDS-Datenbank, einer Amazon-Aurora-Datenbank oder Amazon S3 mit Daten in Ihrer Amazon-Redshift-Datenbank verknüpfen. Mit Amazon Redshift können Sie Betriebsdaten direkt abfragen (ohne sie zu verschieben), Transformationen anwenden und Daten in Ihre Redshift-Tabellen einfügen. Ein Teil der Berechnung für Verbundabfragen wird an die Remote-Datenquellen verteilt.

Um Verbundabfragen auszuführen, stellt Amazon Redshift zunächst eine Verbindung zur Remote-Datenquelle her. Amazon Redshift ruft dann Metadaten zu den Tabellen in der Remote-Datenquelle ab, gibt Abfragen aus und ruft dann die Ergebniszeilen ab. Amazon Redshift verteilt die Ergebniszeilen dann zur weiteren Verarbeitung an Amazon-Redshift-Rechenknoten.

Weitere Informationen zum Einrichten der Umgebung für Verbundabfragen finden Sie in den folgenden Themen im Datenbankentwicklerhandbuch zu Amazon Redshift:

- [Erste Schritte mit der Verwendung von Verbundabfragen an PostgreSQL](#)
- [Erste Schritte beim Verwenden von Verbundabfragen für MySQL](#)

## Zugriff auf Daten in anderen Amazon-Redshift-Clustern

Erfahren Sie, wie Sie mithilfe von Amazon-Redshift-Datenfreigabe Live-Daten sicher und einfach über Amazon-Redshift-Cluster oder AWS-Konten hinweg für Lesezwecke freigeben. Sie profitieren von sofortigem, granulearem und leistungsstarkem Zugriff auf Daten in Amazon-Redshift-Clustern, ohne diese manuell zu kopieren oder zu verschieben. Ihren Benutzer werden aktuelle und konsistente Informationen angezeigt, wenn diese in Amazon-Redshift-Clustern aktualisiert werden. Sie können Daten auf verschiedenen Ebenen freigeben, z. B. Datenbanken, Schemata, Tabellen, Ansichten (einschließlich regulärer Ansichten, Ansichten mit späterer Bindung und materialisierter Ansichten) sowie benutzerdefinierten SQL-Funktionen (UDFs).

Die Amazon-Redshift-Datenfreigabe ist besonders für folgende Anwendungsfälle nützlich:

- Zentralisierung geschäftskritischer Workloads – Verwenden Sie einen zentralen Extract, Transform, Load (ETL)-Cluster, der Daten mit mehreren Business Intelligence (BI)- oder Analyse-Clustern gemeinsam verwendet. Dieser Ansatz bietet Lese-Workload-Isolation und Rückbelastung für einzelne Workloads.
- Freigabe von Daten zwischen Umgebungen – Teilen Sie Daten in Entwicklungs-, Test- und Produktionsumgebungen. Sie können die Teamagilität verbessern, indem Sie Daten auf verschiedenen Granularitätsstufen teilen.

Weitere Informationen zur Datenfreigabe finden Sie unter [Erste Schritte mit der Datenfreigabe](#) im Datenbankentwicklerhandbuch zu Amazon Redshift.

## Training von Machine-Learning-Modellen mit Amazon-Redshift-Daten

Mit Amazon Redshift Machine Learning (Amazon Redshift ML) können Sie ein Modell trainieren, indem Sie die Daten an Amazon Redshift bereitstellen. Dann erstellt Amazon Redshift ML Modelle, die Muster in den Eingabedaten erfassen. Sie können diese Modelle dann verwenden, um Prognosen für neue Eingabedaten zu generieren, ohne dass zusätzliche Kosten entstehen. Mithilfe von Amazon Redshift ML können Sie Machine-Learning-Modelle mithilfe von SQL-Anweisungen trainieren und sie in SQL-Abfragen für Prognosen aufrufen. Sie können die Genauigkeit der Prognosen weiter verbessern, indem Sie die Parameter iterativ ändern und Ihre Trainingsdaten verbessern.

Amazon Redshift ML erleichtert SQL-Benutzern das Erstellen, Trainieren und Bereitstellen von Machine-Learning-Modellen mit vertrauten SQL-Befehlen. Mit Amazon Redshift ML können Sie Ihre Daten in Amazon-Redshift-Clustern verwenden, um Modelle mit Amazon SageMaker Autopilot zu trainieren und automatisch das beste Modell zu erhalten. Sie können dann die Modelle lokalisieren und Prognosen innerhalb einer Amazon-Redshift-Datenbank erstellen.

Weitere Informationen zu Amazon Redshift ML finden Sie unter [Erste Schritte mit Amazon Redshift ML](#) im Datenbankentwicklerhandbuch zu Amazon Redshift.

# Amazon Redshift – Konzeptionelle Übersicht

Mit Amazon Redshift Serverless können Sie auf Daten zugreifen und diese analysieren, ohne alle Konfigurationen wie bei einem bereitgestellten Data Warehouse vornehmen zu müssen. Ressourcen werden automatisch bereitgestellt und die Data-Warehouse-Kapazität wird intelligent skaliert, um eine schnelle Leistung selbst für anspruchsvollste und unvorhersehbare Workloads zu erzielen. Es fallen keine Kosten an, wenn das Data Warehouse inaktiv ist, Sie zahlen also nur für das, was Sie tatsächlich nutzen. Sie können Daten laden und sofort mit der Abfrage beginnen. Hierfür können Sie Amazon Redshift Query Editor v2 oder Ihr bevorzugtes Business Intelligence (BI)-Tool nutzen. Profitieren Sie von einem optimalen Preis-Leistungs-Verhältnis und nutzen Sie vertraute SQL-Funktionen in einer benutzerfreundlichen Umgebung ohne Verwaltungsaufwand.

Wenn Sie Amazon Redshift zum ersten Mal verwenden, empfehlen wir Ihnen, zunächst die folgenden Abschnitte zu lesen:

- [Übersicht über die Funktionen von Amazon Redshift Serverless](#) – Unter diesem Thema finden Sie eine Übersicht über Amazon Redshift Serverless und seinen wichtigsten Funktionen.
- [Service-Merkmale und Preise](#) – Auf dieser Produktdetailseite erfahren Sie mehr zu den Merkmalen und Preisen von Amazon Redshift Serverless.
- [Erste Schritte mit Amazon Redshift Serverless](#) – Unter diesem Thema erfahren Sie, wie Sie ein Amazon Redshift Serverless Data Warehouse erstellen und mit dem Abfragen von Daten mithilfe von Query Editor v2 beginnen können.

Wenn Sie Ihre Amazon-Redshift-Ressourcen lieber manuell verwalten möchten, können Sie bereitgestellte Cluster für Ihre Datenabfrageanforderungen erstellen. Weitere Informationen finden Sie unter [Amazon-Redshift-Cluster](#).

Wenn Ihre Organisation berechtigt ist und Ihr Cluster in einer AWS-Region erstellt wird, in der Amazon Redshift Serverless nicht verfügbar ist, können Sie möglicherweise im Rahmen des kostenlosen Testprogramms von Amazon Redshift einen Cluster erstellen. Wählen Sie entweder Produktion oder Kostenlose Testversion als Antwort auf die Frage: Wofür möchten Sie diesen Cluster verwenden? Wenn Sie Kostenlose Testversion auswählen, erstellen Sie eine Konfiguration mit dem Knotentyp dc2.large. Weitere Informationen zur Auswahl einer kostenlosen Testversion finden Sie unter [Kostenloses Testprogramm für Amazon Redshift](#). Eine Liste der AWS-Regionen, in denen Amazon Redshift Serverless verfügbar ist, finden Sie in der Übersicht über die Endpunkte für die [Redshift-Serverless-API](#) in der Allgemeine Amazon Web Services-Referenz..

Im Folgenden sind einige wichtige Konzepte von Amazon Redshift Serverless aufgeführt.

- **Namespace** – Eine Sammlung von Datenbankobjekten und Benutzern. In Namespaces sind alle Ressourcen zusammengefasst, die Sie in Amazon Redshift Serverless verwenden, wie Schemas, Tabellen, Benutzer, Datashares und Snapshots.
- **Arbeitsgruppe** – Eine Sammlung von Rechenressourcen. In Arbeitsgruppen sind Rechenressourcen enthalten, die Amazon Redshift Serverless zur Ausführung von Datenverarbeitungsaufgaben verwendet. Einige Beispiele für solche Ressourcen sind Redshift Processing Units (RPU), Sicherheitsgruppen und Nutzungslimits. Arbeitsgruppen verfügen über Netzwerk- und Sicherheitseinstellungen, die Sie über die Amazon-Redshift-Serverless-Konsole, die AWS Command Line Interface oder die Amazon-Redshift-Serverless-APIs konfigurieren können.

Weitere Informationen zum Konfigurieren von Namespace- und Arbeitsgruppenressourcen finden Sie unter [Arbeiten mit Namespaces](#) und [Arbeiten mit Arbeitsgruppen](#).

Im Folgenden sind einige wichtige Konzepte im Zusammenhang mit von Amazon Redshift bereitgestellten Clustern aufgeführt:

- **Cluster** – Die zentrale Infrastrukturkomponente eines Amazon-Redshift-Data-Warehouse ist ein Cluster.

Ein Cluster besteht aus einem oder mehreren Datenverarbeitungsknoten. Die Datenverarbeitungsknoten führen den kompilierten Code aus.

Wird ein Cluster mit zwei oder mehr Datenverarbeitungsknoten bereitgestellt, koordiniert ein zusätzlicher Führungsknoten die Datenverarbeitungsknoten. Der Führungsknoten übernimmt die externe Kommunikation mit Anwendungen, wie Business-Intelligence-Tools und Abfrage-Editoren. Ihre Client-Anwendung interagiert nur mit dem Führungsknoten direkt. Die Datenverarbeitungsknoten sind für externe Anwendungen transparent.

- **Datenbank** – Ein Cluster enthält eine oder mehrere Datenbanken.

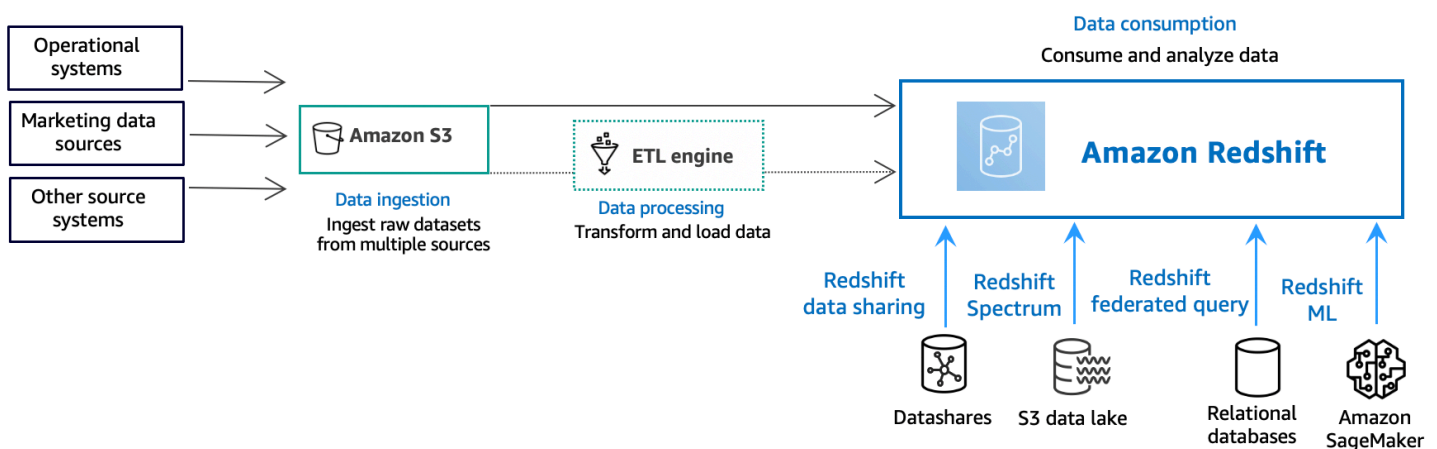
Benutzerdaten werden in einer oder mehreren Datenbanken auf den Datenverarbeitungsknoten gespeichert. Ihr SQL-Client kommuniziert mit dem Führungsknoten, der wiederum die Abfrageausführung mit den Datenverarbeitungsknoten koordiniert. Weitere Informationen zu Datenverarbeitungs- und Führungsknoten finden Sie unter [Data-Warehouse-Systemarchitektur](#). Innerhalb einer Datenbank sind Benutzerdaten in einem Schema oder mehreren Schemata organisiert.



Amazon Redshift ist ein relationales Datenbankmanagementsystem (RDBMS) und ist mit anderen RDBMS-Anwendungen kompatibel. Amazon Redshift stellt dieselben Funktionen wie ein typisches RDBMS bereit, einschließlich Funktionen zur Online-Transaktionsverarbeitung (Online Transaction Processing, OLTP), wie das Einfügen und Löschen von Daten. Amazon Redshift ist auch für leistungsfähige Batchanalysen und Berichterstattung von Datensätzen optimiert.

Im Folgenden finden Sie eine Beschreibung des typischen Datenverarbeitungsablaufs in Amazon Redshift sowie Beschreibungen verschiedener Teile im Ablauf. Weitere Informationen zur Amazon-Redshift-Systemarchitektur finden Sie unter [Architektur des Data-Warehouse-Systems](#).

Das folgende Diagramm zeigt einen typischen Datenverarbeitungsablauf in Amazon Redshift.



Ein Amazon-Redshift-Data-Warehouse ist eine Abfrage- und Verwaltungssystem der Enterprise-Klasse für relationale Datenbanken. Amazon Redshift unterstützt Client-Verbindungen mit vielen Arten von Anwendungen, einschließlich Business Intelligence (BI), Berichterstellung, Daten und Analysetools. Bei Analyseabfragen werden große Datenmengen in mehrphasigen Operationen abgerufen, verglichen und bewertet, um ein Endergebnis zurückzugeben.

In der Ebene der Datenerfassung laden verschiedene Arten von Datenquellen kontinuierlich strukturierte, halbstrukturierte oder unstrukturierte Daten in die Datenspeicher-Ebene hoch. Dieser Datenspeicherbereich dient als Staging-Bereich, der Daten in verschiedenen Zuständen der Nutzungsbereitschaft speichert. Ein Beispiel für einen solchen Speicher ist ein Amazon Simple Storage Service (Amazon S3)-Bucket.

In der optionalen Ebene Datenverarbeitung durchlaufen die Quelldaten die Vorverarbeitung, Validierung und Transformation über Extract, Transform, Load (ETL)-oder Extract, Load, Transform

(ELT)-Pipelines. Diese Rohdatensätze werden dann mithilfe von ETL-Operationen verfeinert. Ein Beispiel für eine ETL-Engine ist AWS Glue.

In der Ebene Datennutzung werden Daten in Ihren Amazon-Redshift-Cluster geladen, wo Sie Analyse-Workloads ausführen können.

Beispiele für Analyse-Workloads finden Sie unter [Abfragen von externen Datenquellen](#).

## Weitere Ressourcen

Wenn Sie mehr über Amazon Redshift Serverless erfahren möchten, empfehlen wir Ihnen, Ihr Wissen über die in diesem Handbuch vorgestellten Konzepte unter Verwendung der folgenden Ressourcen zu Amazon Redshift zu vertiefen:

- **Feature-Videos:** Diese Videos helfen Ihnen, mehr über die Funktionen von Amazon Redshift zu erfahren.
  - Um eine allgemeine Vorstellung von Amazon Redshift Serverless zu erhalten, sehen Sie sich das folgende Video an. [Amazon Redshift Serverless Explained in 90 Seconds](#) (Amazon Redshift Serverless in 90 Sekunden erklärt).
  - Wenn Sie wissen möchten, wie Sie ein Serverless Data Warehouse einrichten und mit dem Abfragen von Daten beginnen können, sehen Sie sich das folgende Video an. [Getting Started with Amazon Redshift Serverless](#) (Erste Schritte mit Amazon Redshift Serverless)
- [Amazon-Redshift-Verwaltungshandbuch](#): Dieser Leitfaden baut auf dem Handbuch Erste Schritte mit Amazon Redshift auf. Sie erhalten darin detaillierte Informationen über die Konzepte und Aufgaben für die Erstellung, Verwaltung und Überwachung von Clustern, die von Amazon Redshift Serverless und Amazon Redshift bereitgestellt werden.
- [Datenbankentwicklerhandbuch zu Amazon Redshift](#): Dieses Handbuch baut ebenfalls auf dem Handbuch Erste Schritte mit Amazon Redshift auf. Es richtet sich an Datenbankentwickler und vermittelt fundierte Kenntnisse auf den Gebieten Entwurf, Entwicklung, Abfrage und Verwaltung von Datenbanken in einem Data Warehouse.
  - [SQL-Referenz](#): In diesem Thema werden SQL-Befehle und Funktionsreferenzen für Amazon Redshift beschrieben.
  - [Systemtabellen und Ansichten](#): In diesem Thema werden Systemtabellen und Ansichten für Amazon Redshift beschrieben.
- **Tutorials für Amazon Redshift:** In diesem Thema werden Tutorials zu den Funktionen von Amazon Redshift angezeigt.
  - [So laden Sie Daten aus Amazon S3](#): In diesem Tutorial wird beschrieben, wie Sie Daten aus Datendateien in einem Amazon-S3-Bucket in Ihre Amazon-Redshift-Datenbanktabellen laden.
  - [Erste Schritte mit der Datenfreigabe](#): In diesem Abschnitt wird beschrieben, wie Sie Daten freigeben und auf Daten in anderen Amazon-Redshift-Clustern zugreifen können.
  - [Verwendung von Geo-SQL-Funktionen mit Amazon Redshift](#): In diesem Tutorial wird gezeigt, wie Sie einige der Geo-SQL-Funktionen mit Amazon Redshift verwenden.

- [Abfragen verschachtelter Daten mit Amazon Redshift Spectrum](#): In diesem Tutorial wird beschrieben, wie Sie Redshift Spectrum verwenden, um verschachtelte Daten in den Dateiformaten Parquet, ORC, JSON und Ion mit externen Tabellen abzufragen.
- [Konfigurieren von manuellen Workload-Management \(WLM\)-Warteschlangen](#): In diesem Tutorial wird beschrieben, wie Sie das manuelle Workload-Management (WLM) in Amazon Redshift konfigurieren.
- [Erste Schritte mit Amazon Redshift ML](#): In diesem Abschnitt wird beschrieben, wie Benutzer unter Verwendung von vertrauten SQL-Befehlen Machine-Learning-Modelle erstellen, trainieren und bereitstellen können.
- [Neuerungen](#): Diese Webseite listet neue Funktionen von Amazon Redshift und Produktaktualisierungen auf.

# Dokumentverlauf

## Note

Eine Beschreibung der neuen Funktionen in Amazon Redshift finden Sie unter [Was ist neu?](#).

In der folgenden Tabelle werden die wichtigen Dokumentationsänderungen am Handbuch Erste Schritte mit Amazon Redshift beschrieben.

Änderung	Beschreibung	Datum der Veröffentlichung
Aktualisierung der Dokumentation	Das Handbuch wurde aktualisiert, um neue Abschnitte über die ersten Schritte mit allgemeinen Datenbankaufgaben, das Abfragen Ihres Data Lake, das Abfragen von Daten auf entfernten Quellen, das Freigeben von Daten und das Training von Machine-Learning-Modellen mit Amazon-Redshift-Daten zu enthalten.	30. Juni 2021
Neues Feature	Aktualisierung des Handbuchs, um das Verfahren des neuen Beispielladevorgangs zu beschreiben.	4. Juni 2021
Aktualisierung der Dokumentation	Die ursprüngliche Amazon-Redshift-Konsole wurde aus dem Handbuch entfernt und der Ablauf der Schritte wurde verbessert.	14. August 2020
New console	Das Handbuch enthält nun eine Beschreibung der neuen Amazon-Redshift-Konsole.	11. November 2019
Neues Feature	Aktualisierung des Handbuchs, um das Verfahren für schnelle Clusterstarts zu beschreiben.	10. August 2018
Neues Feature	Das Handbuch enthält nun Informationen zum Starten von Clustern über das Amazon-Redshift-Dashboard.	28. Juli 2015

Änderung	Beschreibung	Datum der Veröffentlichung
Neues Feature	Das Handbuch enthält nun Informationen zur Verwendung von neuen Knotentypnamen.	9. Juni 2015
Aktualisierung der Dokumentation	Die Dokumentation enthält aktualisierte Screenshots und Verfahren zur Konfiguration der VPC-Sicherheitsgruppen.	30. April 2015
Aktualisierung der Dokumentation	Die Dokumentation enthält aktualisierte Screenshots und beschreibt, wie sich die aktuelle Konsole und die Screenshots aufeinander abstimmen lassen.	12. November 2014
Aktualisierung der Dokumentation	Zur besseren Auffindbarkeit wurden die Informationen zum Laden von Daten von Amazon S3 in einen separaten Abschnitt verschoben und der Abschnitt mit den nächsten Schritten wurde in den letzten Schritt integriert.	13. Mai 2014
Aktualisierung der Dokumentation	Die Willkommenseite wurde entfernt; ihr Inhalt befindet sich nun auf der Hauptseite der Seite "Erste Schritte".	14. März 2014
Aktualisierung der Dokumentation	Bei dieser Dokumentation handelt es sich um eine neue Version des Handbuchs Erste Schritte mit Amazon Redshift, die Kundenfeedback und Service-Updates berücksichtigt.	14. März 2014
Neues Handbuch	Dies ist die erste Version des Handbuchs Erste Schritte mit Amazon Redshift.	14. Februar 2013

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.